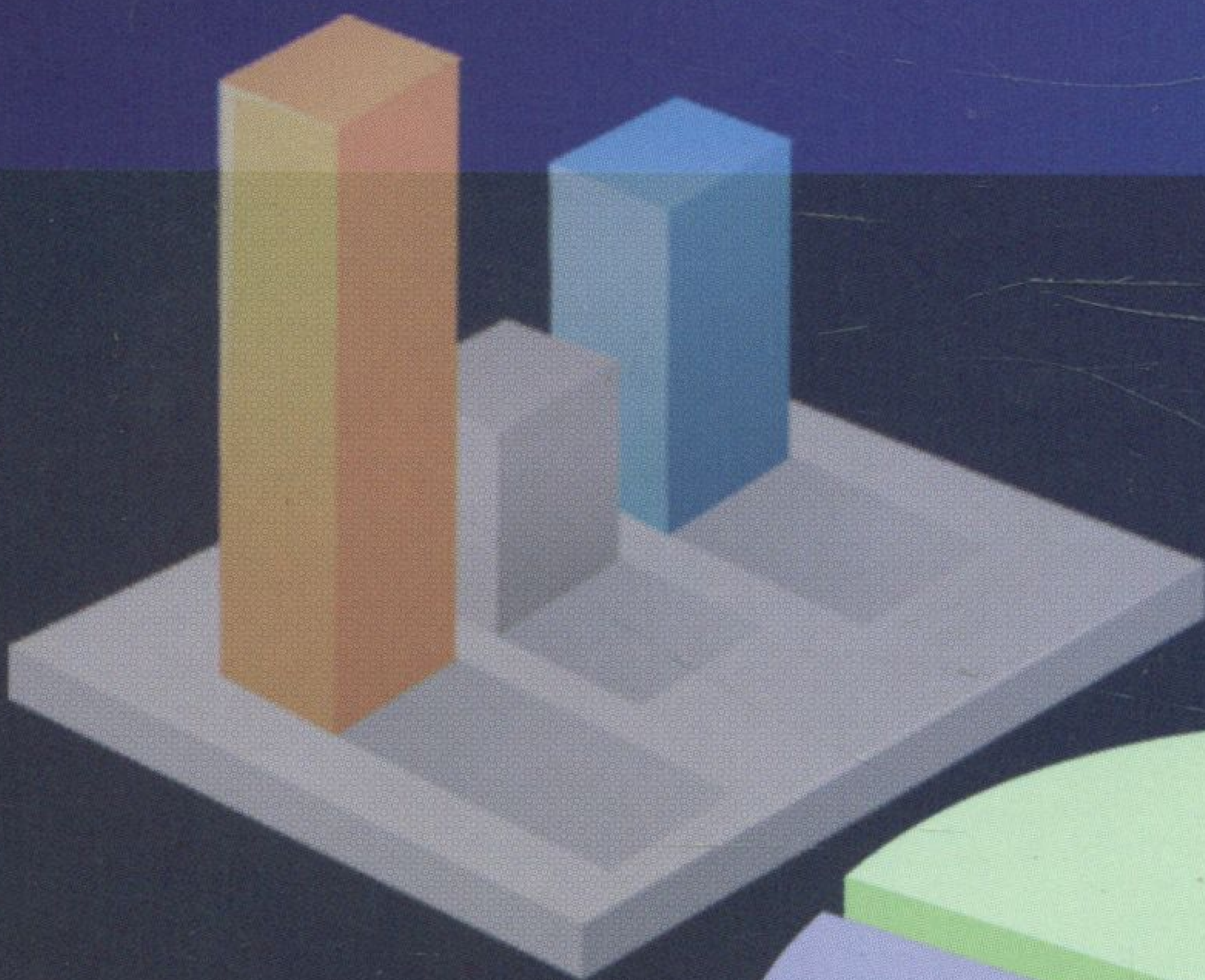
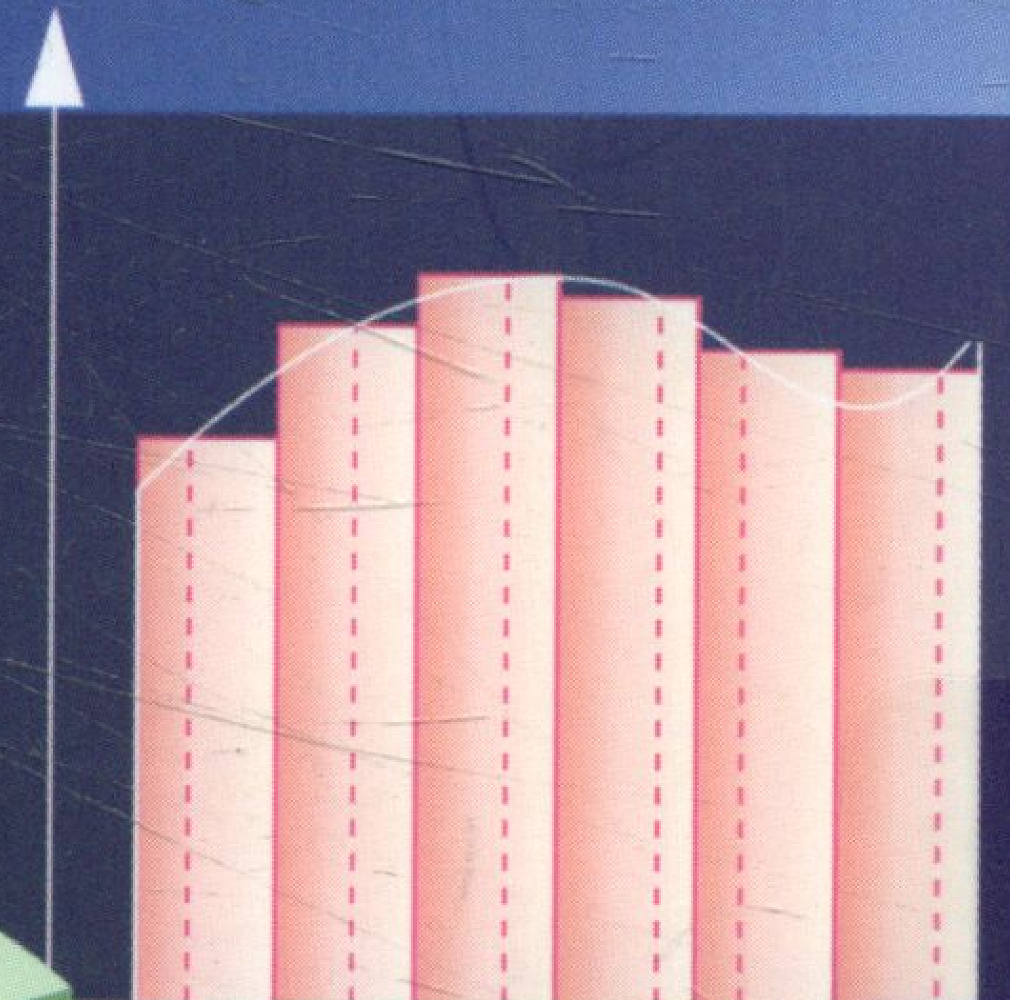
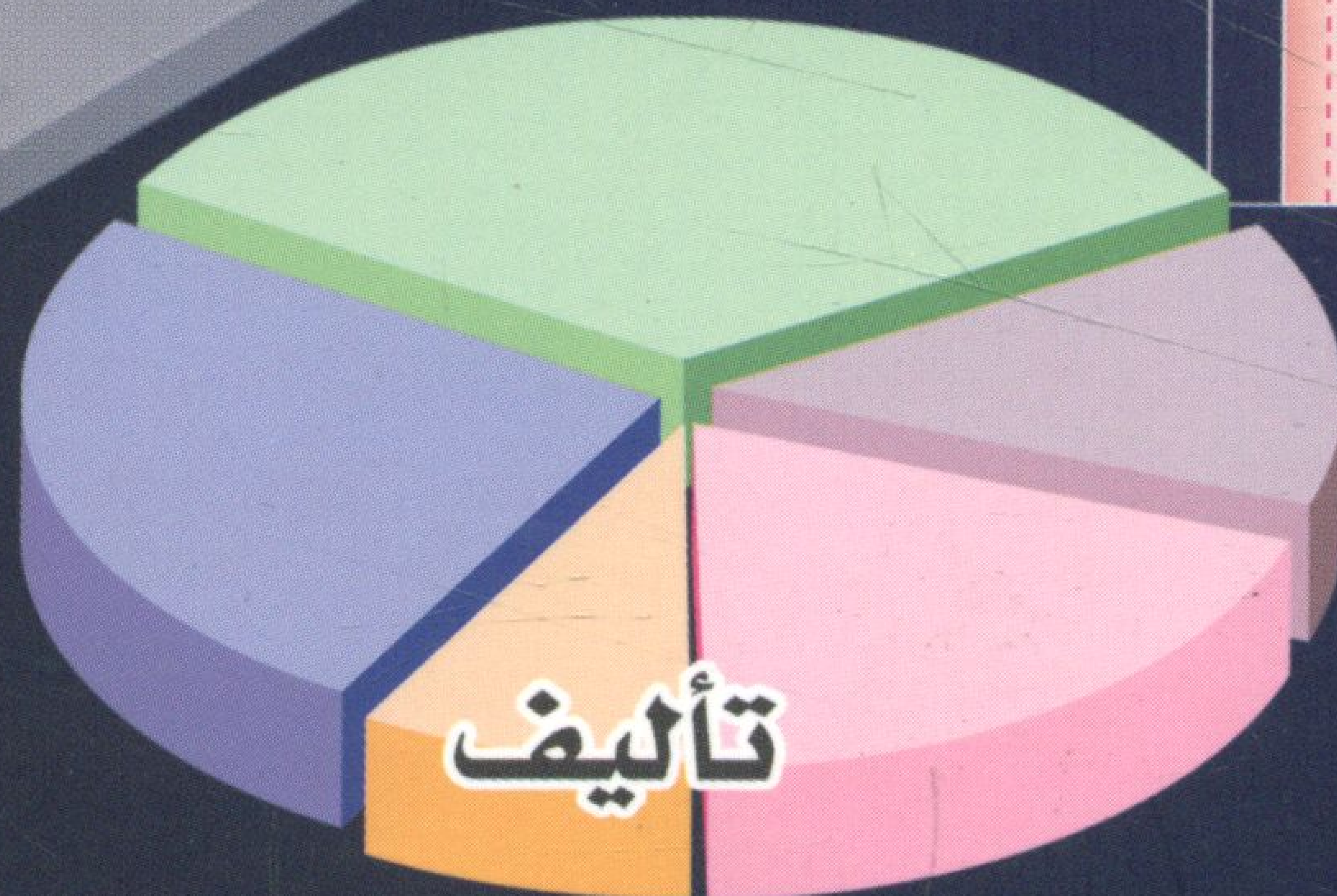


جامعة سلمان بن عبدالعزيز
Salman bin Abdulaziz University

مبادئ الإحصاء والاحتمالات وتطبيقاتها باستخدام SPSS







تأليف

د. عبد الرحمن بن إبراهيم محمد الخضيري

د. عبد الفتاح مصطفى محمد السيد

مبادئ الإحصاء والاحتمالات
وتطبيقاتها باستخدام SPSS

الدكتور

عبدالرحمن بن إبراهيم محمد الخضيري

أستاذ الإحصاء الرياضي المشارك

قسم الإحصاء وبحوث العمليات –كلية العلوم

جامعة الملك سعود

الدكتور

عبدالفتاح مصطفى محمد السيد

أستاذ الإحصاء الرياضي المشارك

قسم الرياضيات – كلية العلوم والدراسات الإنسانية

جامعة سلمان بن عبدالعزيز

ⓒ جامعة سلمان بن عبدالعزيز، ١٤٣٤هـ/ (٢٠١٣م)

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

الخضيري، عبدالرحمن إبراهيم

مبادئ الاحصاء والاحتمالات وتطبيقاتها باستخدام SPSS./ عبدالرحمن ابراهيم الخضيري؛ عبدالفتاح مصطفي السيد. - الخرج، ١٤٣٤هـ

٥٠٠ ص؛ ٢٨X٢١سم

ردمك: ٦ - ٠ - ٩٠٤٤٤ - ٦٠٣ - ٩٧٨

١- التحليل الاحصائي أ. السيد، عبدالفتاح مصطفى (مؤلف مشارك) ب. العنوان

ديوي ٥١٩,٥ ١٤٣٤/٤١٠٨

رقم الايداع: ١٤٣٤/٤١٠٨

ردمك: ٦ - ٠ - ٩٠٤٤٤ - ٦٠٣ - ٩٧٨

حكمت هذا الكتاب لجنة متخصصة شكلها المجلس العلمي، وقد وافق المجلس على طباعته ونشره في اجتماعه الثالث عشر للعام الدراسي١٤٣٢/١٤٣٣هـ المعقود بتاريخ ١٢/ ٧/ ١٤٣٣هـ الموافق ٢/ ٦/ ٢٠١٢م.

بسم الله الرحمن الرحيم

مقدمة

INTRODUCTION

يعدّ هذا الكتاب أساساً لعلم الإحصاء والاحتمالات وأيضاً مدخلاً للاستدلال الإحصائي، كما يعد مرجعاً مهماً لكل دارسي الإحصاء بجميع تخصصاتهم العلمية والهندسية حيث يحتوي هذا الكتاب على المبادئ الأساسية لعلم الإحصاء وكيفية وصف وعرض البيانات الإحصائية وتمثيلها بيانياً في الحالات المختلفة للبيانات؛ سواء كانت بيانات لمتغير واحد فقط سواء كان وصفياً أو كمياً. كما يستعرض المفاهيم الأساسية لعلم الاحتمالات التي هي أساس يجب أن يتعرض لها كل من يدرس الاحتمالات من تعريفات ونظريات ذات أهمية في هذا العلم إضافة إلى دراسة الاحتمال الشرطي وتطبيقاته. كما تطرق هذا الكتاب لبعض التوزيعات الاحتمالية للمتغيرات العشوائية المنفصلة وأيضا المتصلة وتطبيقاتها. وأخيراً تم عرض مقدمة أساسية للاستدلال الإحصائي ابتداءً من توزيعات المعاينة ثم الانتقال لتقدير معالم المجتمع الإحصائي المجهولة أو انتهاءً باختبار ادعاء (فرض) حول معالم المجتمع الإحصائي.

ومما يميز هذا الكتاب أنه يهتم بإجراء بعض التطبيقات الإحصائية باستخدام برنامج العرض والتحليل الإحصائي SPSS النسخة السابعة عشر وقد قمنا بالتركيز على برنامج SPSS لما له من أهمية في إجراء التطبيقات الإحصائية وسهولة التعامل معه.

ويمكن تقسيم هذا الكتاب إلي أحد عشر فصلاً كما يلي:

الفصل الأول بعنوان بعض المفاهيم الأساسية

يحوي هذا الفصل المفاهيم الأساسية في علم الإحصاء من التعريفات والمصطلحات التي سنحتاج لها في الفصول التالية.

الفصل الثاني بعنوان تبويب البيانات وتمثيلها بيانياً

ويحتوى هذا الفصل على كيفية جدولة البيانات الوصفية والبيانات الكمية وأيضا تمثيلها بيانياً وينتهي هذا الفصل ببعض التطبيقات باستخدام برنامج SPSS.

الفصل الثالث بعنوان مقاييس النزعة المركزية

ويحتوي هذا الفصل على بعض المقاييس الإحصائية للمتغيرات الكمية والتي تختص بالنزعة المركزية وقمنا بإجراء بعض التطبيقات لتعيين مقاييس النزعة المركزية باستخدام برنامج SPSS.

الفصل الرابع بعنوان مقاييس التشتت

ويشمل هذا الفصل على بعض المقاييس التي تستخدم في دراسة تشتت البيانات والتي نستطيع من خلالها التعرف على مدى تجانس البيانات ومقارنه البيانات بعضها ببعض سواء لمتغير واحد أو أكثر من متغير وتم التعرض لبعض المقاييس التي تستخدم في دراسة طبيعة منحنى البيانات وتم تقديم بعض التطبيقات للتعرف على كيفية تعيين مقاييس التشتت باستخدام برنامج SPSS.

الفصل الخامس بعنوان الوصف الإحصائي لمتغيرين

ويشمل هذا الفصل بعض الطرق التي تستخدم في وصف متغيرين والانحدار الخطي البسيط، وتم عرض معامل الارتباط لدراسة طبيعة العلاقة بين المتغيرين. واخيراً تم تقديم بعض التطبيقات للتعرف على كيفية استخدام برنامج SPSS في الوصف الإحصائي لمتغيرين.

الفصل السادس بعنوان الاحتمالات

لأهمية الاحتمالات وتطبيقاتها فإننا تعرضنا خلال هذا الفصل للمبادئ الأساسية للاحتمالات، ومفهوم الاحتمال الشرطي وتطبيقاته لما له من أهمية.

الفصل السابع بعنوان المتغيرات العشوائية

خلال الفصل السادس تم التعرض لمفهوم الحادثة وتعيين احتمالها، ولكن هذا غير شائع الاستخدام، فكثيراً ما نتعامل مع المتغيرات العشوائية وليست الحوادث لذا فان هذا الفصل يتعرض لمفهوم المتغيرات العشوائية وأنواعها ووصفها ودراسة بعض المفاهيم الخاصة بالمتغيرات العشوائية.

الفصل الثامن بعنوان بعض التوزيعات الاحتمالية المنفصلة

يعرض هذا الفصل بعض التوزيعات الاحتمالية للمتغيرات العشوائية المنفصلة والتي يطلق عليها التوزيعات الاحتمالية المنفصلة.

الفصل التاسع بعنوان بعض التوزيعات الاحتمالية المتصلة

بعد التعرض في الفصل الثامن لبعض التوزيعات الاحتمالية المنفصلة كان لابد في هذا الفصل من التعرض لبعض التوزيعات الاحتمالية المتصلة والتي لها تطبيقات عملية في عدد من المجالات الهندسية والطبية والعلمية.

الفصل العاشر بعنوان مبادئ توزيعات المعاينة

في هذا الفصل تم التعرض لكيفية جمع البيانات لدراسة إحصائية باستخدام طريقة المسح الشامل أو طريقة العينة الإحصائية. وأيضا قمنا بعرض التوزيعات الاحتمالية للإحصاءات الخاصة بالعينة.

الفصل الحادي عشر بعنوان مبادئ الاستدلال الإحصائي

وفى نهاية هذا الكتاب كان لابد من الربط بين الإحصاء والاحتمالات من الناحية النظرية والتطبيقات العملية فى تقديم نبذة مختصرة عن مبادئ الاستدلال الاحصائى. وقدمنا للقارئ تطبيقات عملية لفترات الثقة واختبارات الفروض باستخدام برنامج SPSS.

أيضا تم تقديم بعض الأسئلة والتمارين في نهاية كل فصل والتي يمكن من خلالها تقويم ما تم التعرض له خلال الفصل. ويتبع الفصول السابقة الجداول الإحصائية التي سوف نحتاج إليها خلال فصول الكتاب ثم قائمة بالمراجع ثم كشاف الموضوعات.

كما أننا نضع هذا الكتاب بين أيدي الإخوة الزملاء وطلاب العلم والمعرفة وسنكون شاكرين لهم إذا تفضلوا بتزويدنا بأي نقص أو اقتراح بَناء يساعدنا على تحسينه مستقبلاً حتى يكون أكثر نفعاً.

ويخص كل من المؤلفين زوجه وأبناءه بالشكر الجزيل لاستقطاع بعض من وقتهم لإعداد هذا الكتاب.

وأخيراً نسأل الله تعالى أن يجعل هذا العمل خالصاً لوجهه الكريم وأن ينفع به الإسلام والمسلمين، والله من وراء القصد وهو الهادي إلى سواء السبيل.

المؤلفان

عبدالرحمن إبراهيم الخضيري

عبدالفتاح مصطفى السيد

المحتويات

الصفحة

الفصل الثالث: مقاييس النزعة المركزية

الفصل الرابع: مقاييس التشتت

الفصل الخامس: الوصف الإحصائي لمتغيرين

الفصل السادس: الاحتمالات

الفصل السابع: المتغيرات العشوائية

الفصل الثامن: بعض التوزيعات الاحتمالية المنفصلة

الفصل التاسع: بعض التوزيعات الاحتمالية المتصلة

الفصل العاشر:مبادئ توزيعات المعاينة

الفصل الحادي عشر: مبادئ الاستدلال الإحصائي

الفصل الأول

بعض المفاهيم الأساسية

SOME BASIC CONCEPTS

١–١ مقدمة

الإحصاء كلمة كثيراً ما نسمعها في وسائل الأعلام ونقرأها في الجرائد والمجلات عندما يراد التعبير عن تطوير ما في مجال معين. ويتصور كثير من الناس أن الإحصاء عبارة عن مجموعة من الأرقام أو مجموعة من الرسومات والمنحنيات تمثل تطور ظاهرة معينة. ربما كانت هذه هي الحالة في الماضي عندما كانت مهمة الإحصائي جمع البيانات وتنظيمها في جداول ثم عرضها بيانياً. أما الإحصائي اليوم فانه يقوم بأعمال علمية ومهنية مشوقة ومعقدة. فهو يساعد في تصميم التجارب ويقوم بتحليل البيانات وتفسيرها واتخاذ القرار المناسب على ضوء ما توصل إليه من نتائج.

ترتبط حياتنا اليومية بالإحصاء في العديد من المجالات سواء على مستوى الأفراد أو الجماعات فالعديد من الظواهر أو المتغيرات التي نعيشها سواء ما كان منها ذا طابع اقتصادي أو اجتماعي أو تعليمي أو تربوي أو صحي أو نفسي.... إلى غير ذلك من الظواهر، تتم دراستها إما عن طريق مشاهدة تطوراتها وتفاعلاتها مع بعضها البعض أو تخضع لأسلوب التجريب للتعرف على طبيعتها وتأثيرها بعضها ببعض. والسبيل الأمثل (إن لم يكن الأوحد) إلى دراسة هذه الظواهر وتفسيرها والتنبؤ بما ستكون عليه في المستقبل هو الأسلوب الإحصائي أو أسلوب دراسة المتغيرات.

تُفيد الطرق الإحصائية الحديثة في فحص ودراسة أنواع كثيرة من المشاكل المختلفة والمتعددة وعلى سبيل المثال لا الحصر سنذكر بعض المشاكل التي تستطيع الطرق الإحصائية المساعدة في إيجاد حل لها:

- كيف يمكن اختبار تأثير وفاعلية دواء معين؟
- كيف نقيس تأثير تدخين السجائر في زيادة احتمال الإصابة بمرض السرطان؟
- كيف يمكن المقارنة بين عدة أنواع من الأسمدة لاختيار الأفضل منها؟
- ما تأثير السمنة على طول حياة الإنسان؟
- كيف يمكن الرقابة على الإنتاج بحيث يمكن معرفة الخلل فور حدوثه؟
- كيف يمكن تقدير متوسط دخل الأسرة في دولة ما؟
- كيف نتنبأ بإنتاج أحد المحاصيل في إحدى السنوات؟
- كيف نقدر نسبة التالف في إنتاج مصنع معين؟
- كيف نقيس العلاقة بين عمر الزوج وعمر الزوجة؟

هذه بعض الأسئلة لأمثلة كثيرة تستطيع الطرق الإحصائية المساعدة في الإجابة عليها وعملياً في كل يوم يقابل الإحصائيين مشاكل جديدة تحتاج إلى معالجة إما بطرق جديدة أو بإجراء تغيير ما في طريقة معروفة. وقد يواجه الإحصائي عند إجراء أي دراسة إحصائية بعض المشكلات في الحسابات العددية التي يقوم بها بعد جمع البيانات إما في عملية وصف البيانات وتمثيلها بيانياً أو تحليلها إحصائيا للوصول إلى نتائج إحصائية يتم اتخاذ قرار بناء عليها. لكن مع تطور الحاسبات وعلم البرمجة أصبحت تلك المشكلة من أبسط الأجزاء بالنسبة للإحصائي. فتطور الكمبيوتر أدى إلى وجود ما يسمى بالبرامج الجاهزة التي تم توجيهها للمجالات المختلفة وقد حظي الإحصاء باهتمام كبير من قبل المبرمجين. وتم تصميم بعض البرامج الجاهزة التي تخدم الإحصائي في إجراء الحسابات العددية وتلك البرامج في تطور مستمر نتيجة لتطور علم الإحصاء وفروعه المختلفة. وسوف نتعرض خلال هذا الفصل لنبذة مختصرة عن برنامج من أهم البرامج الإحصائية، وهو برنامج SPSS

١-٢ تعريف الإحصاء

الإحصاء هو العلم الذي يبحث في:

(أ) جمع البيانات والمعلومات والحقائق الخاصة بمختلف الظواهر وتسجيلها في صورة رقمية وتصنيفها في جداول منظمة وتمثيلها بيانياً.

(ب) تحليل البيانات واستخلاص النتائج منها واتخاذ القرارات المناسبة.

(ج) مقارنة البيانات ودراسة العلاقات بينها واستخدامها في تفهم حقيقة الظواهر ومعرفة القوانين التي تسير تبعا لها.

ولا نريد الخوض في ما إذا كان الإحصاء علماً فقط أو هو إلى جانب ذلك فن أو طريقة أو أسلوب أو أداة، بل لعله كل ذلك جميعا، فهو للمتخصصين قاعدة أساسية لانطلاق الفكر عبر النظريات الرياضية والتطبيقات الاقتصادية والاجتماعية والعلمية بحثا عن الحقيقة بصورة علمية منظمة. وهو لغير المتخصصين أداة تنسيق وتنظيم وتحليل بهدف استكشاف الحقيقة بطريقة علمية منظمة. وفيما يلي سنشير باختصار شديد إلى بعض المفاهيم الأساسية التي سوف نحتاج إليها فيما بعد.

١-٣ أنواع الإحصاء

يمكن تقسيم علم الإحصاء إلى نوعين أساسين وهما:

(أ) الإحصاء الوصفي (Descriptive Statistics)

يتكون الإحصاء الوصفي من مجموعة الأساليب التي تعنى بجمع البيانات وتنظيمها وتلخيصها وعرضها بطريقة واضحة في صورة جداول وأشكال بيانية وحساب المقاييس الإحصائية المختلفة لها. وعند إجراء دراسة إحصائية لمشكلة ما فإنه يجب علينا أولا تحديد جميع المفردات أو الوحدات التي تتعلق بهذه المشكلة حتى نقوم بجمع البيانات عنها من هذه المفردات.

(ب) الإحصاء الاستدلالي (Inferential Statistics)

الإحصاء الاستدلالي يُعني باختيار جزء من المجتمع يسمى عينة بطريقة علمية مناسبة، بغرض استخدام بيانات هذه العينة في التوصل إلى نتائج، يمكن تعميمها على مجتمع الدراسة، ومن ثم يهتم الاستدلال الإحصائي بموضوعين هما التقدير- اختبارات الفروض.

يتركز اهتمامنا في الدراسات الإحصائية على بعض الظواهر أو المتغيرات المتعلقة بالمجتمع محل الدراسة. والقيم التي تأخذها هذه المتغيرات تختلف من مفردة لأخرى، هذا الاختلاف يرجع إلى عوامل عشوائية ولذلك من الصعب التنبؤ بقيمة هذه المتغيرات ومثل هذه المتغيرات تسمى متغيرات عشوائية.

ويستخدم الإحصاء الوصفي والإحصاء الاستدلالي في جميع المجالات وهما متمم لبعضهما البعض، فعند إجراء أي دراسة إحصائية يتم أولا جمع البيانات عن تلك الدراسة وعرض تلك البيانات في صورة سهلة وبسيطة وهذا هو دور الإحصاء الوصفي ثم الاستفادة من تلك البيانات وتحليلها إحصائياً باستخدام الطرق الإحصائية والاستدلال منها على معلومات ونتائج واتخاذ بعض القرارات بناء على النتائج التي تم التوصل إليها، وهذا هو دور الإحصاء الاستدلالي.

وأحيانا يأخذ علم الإحصاء مسماه من طبيعة البيانات والدراسة التي نقوم بها لذا يمكن تصنيف الإحصاء إلي:

- الإحصاء الحيوي
- الإحصاء التربوي
- الإحصاء النفسي
- الإحصاء الاجتماعي
- الإحصاء السكاني
- الإحصاء الرياضي
- الإحصاء الهندسي
- الإحصاء التطبيقي

وهذا يدل على أهمية علم الإحصاء وأهمية تطبيقاته في المجالات المختلفة. ودائما نقول إن علم الإحصاء هو علم قائم بذاته يخدم جميع العلوم.

١-٤ بعض المفاهيم الأساسية

المجتمع (Population):

هو مجموعة كل الأفراد (المفردات) موضع الاهتمام في دراسة معينة. وقد يكون المجتمع محدوداً (finite)، إذا تكون من عدد محدود من المفردات، أو يكون لا نهائياً (infinite) إذا كان يحتوي على عدد غير محدود من المفردات. كما أن المجتمع قد يكون حقيقياً (real) أو افتراضياً (hypothetical). وإذا تم جمع البيانات من جميع مفردات المجتمع فتسمى هذه العملية بالحصر الشامل.

إننا نعيش في عصر يصعب فيه الحكم على صحة الأشياء بدون الأدلة الموضوعية والبحث العلمي الواسع. عصر عمت فيه المشكلات العلمية المختلفة والظواهر المتعددة التي يصعب وصفها بدون دراسات وبحث. ولدراسة أي مشكلة علمية فإننا نحتاج أولاً إلى جمع كل ما يتعلق بتلك المشكلة من بيانات ومعلومات وبناء على هذه المعلومات نتوجه لدراسة تلك المشكلة وتسمى مجموعة العناصر المتعلقة بتلك المشكلة بالمجتمع الإحصائي. ومن هنا نجد إن أي دراسة إحصائية تبدأ دائما بالبيانات الخام التي يتم توفيرها لتلك الدراسة وتلك البيانات يمكن جمعها عن طريق جميع عناصر المجتمع الإحصائي أو عن طريق جزء من ذلك المجتمع الإحصائي.

طريقة المسح الشامل:

الطريقة التي يتم بها جمع البيانات من جميع أفراد المجتمع الإحصائي تسمى بطريقة المسح الشامل فعلى سبيل المثال إذا أردنا دراسة ظاهرة التدخين في مجتمع ما وتجميع البيانات باتباع ظاهرة المسح الشامل فإنه يجب إجراء مقابلات شخصية مع جميع أفراد المجتمع وتوجيه سؤال لكل شخص هل هو مدخن أم لا؟.

أيضاً إذا أردنا تعيين عدد سكان المملكة العربية السعودية باتباع طريقة المسح الشامل فإنه يجب توجيه سؤال لكل أسرة عن عدد أفرادها أو تجميع تلك البيانات من السجلات المدنية مع وجود بعض التحفظات، بأنه قد يوجد بعض الأشخاص غير المسجلين في السجلات المدنية أو ما شابه ذلك. وطريقة المسح الشامل تعطى بيانات دقيقة في حالة المجتمعات الصغيرة لكن إذا كان المجتمع كبيراً فتكون هناك صعوبة ومجهود في جمع البيانات باتباع طريقة المسح الشامل. وأيضا توجد حالات أخرى يصعب فيها إجراء طريقة المسح الشامل منها:

١- فساد عناصر المجتمع نتيجة أخذ المشاهدات من عناصر المجتمع، فعلى سبيل المثال إذا أردنا دراسة أعمار المصابيح التي ينتجها مصنع معين باستخدام طريقة المسح الشامل فإننا سنقوم بإنارة جميع المصابيح التي ينتجها المصنع حتى يتم تلفها(خرابها) وهذا يؤدى إلى حدوث خسارة للمصنع.

٢- غالباً ما يتعذر الوصول لجميع عناصر المجتمع الإحصائي وإجراء الدراسة عليها. فإذا أردنا تعيين نسبة المدخنين بالمملكة العربية السعودية باتباع طريقة المسح الشامل فإنه يجب سؤال كل شخص من سكان المملكة هل هو مدخن أم لا؟ مع العلم أنه يوجد أشخاص تقيم في أماكن يصعب التوصل إليها أو أشخاص خارج المملكة لفترة ما.

٣- إن معظم الدراسات الإحصائية تكون مقيدة بمقدار من التكاليف والزمن والجهد المخصص لإنجازها. لهذا قد لا تسمح لنا تلك الظروف والقيود بإجراء المسح الشامل.

٤- قد تحتاج عملية المسح الشامل إلى عدد كبير من الأشخاص لإتمام جمع البيانات وهذا ينتج عنه أخطاء متعددة نتيجة لوجود فروق فردية بين العاملين على جمع البيانات.

٥- قد نحتاج أحياناً إلى نتائج سريعة لاتخاذ قرار معين في حين أن طريقة المسح الشامل قد تحتاج إلى وقت طويل نسبياً.

٦- قد يكون المجتمع الإحصائي متصلاً كأن تكون عناصر المجتمع غير قابلة للعد، فمثلاً إذا أردنا التنقيب عن البترول في بلداً ما، فباتباع طريقة المسح الشامل يجب التنقيب في جميع الأراضي، وهذا أمر غير ممكن.

ولذلك كان لزاماً إيجاد بديل لطريقة المسح الشامل نتيجة للأسباب السابقة، ودراسة جزء فقط من المجتمع الإحصائي فيما يسمى العينة الإحصائية، ثم باستخدام بعض الطرق الإحصائية يمكن التنبؤ بقيم المجتمع الإحصائي.

العينة (Sample):

هي جزء من مفردات المجتمع يتم اختيارها بحيث تكون ممثلة للمجتمع ككل. وأسلوب العينات شائع الاستعمال عند إجراء الدراسات والبحوث الإحصائية لأن تكاليفه أقل، وبواسطته يمكن الحصول على نتائج سريعة وذات دلالة. وتوجد عدة طرق يتم عن طريقها اختيار العينات من المجتمع الإحصائي وهي:

(أ) طريقة العينة العشوائية البسيطة (Simple Random Sample)

ومن مميزات العينة العشوائية البسيطة أن لكل عنصر من عناصر المجتمع الإحصائي نفس فرصة الاختيار في العينة. والعينة هي أي مجموعة جزئية من المجتمع الإحصائي بحجم معين. ويتم اختيار العينة العشوائية البسيطة عن طريق:

- الجداول العشوائية: وذلك بإعطاء كل عنصر من عناصر المجتمع الإحصائي رقما مسلسلاً من صفر حتى $(N-1)$ حيث N حجم المجتمع الإحصائي ونجعل هذه الأرقام مكونه من نفس عدد الخانات، ومن جدول الأرقام العشوائية نقرأ عمودياً الأرقام المكونة من نفس عدد الخانات، فإذا كان العدد الذي نقرأه من الجدول أحد الأرقام المسلسلة قبلناه كأحد عناصر العينة وإلا رفضناه، ثم ننتقل لقراءة عدد أخر، ونستمر في القراءة كلما انتهينا من عمود انتقلنا إلى العمود الذي يليه حتى نحصل على العينة بالحجم المطلوب.
- البطاقات: وذلك بكتابه أرقام أو أسماء عناصر المجتمع الإحصائي على بطاقات متماثلة من حيث الشكل والحجم، ويتم غلقها بنفس الطريقة ووضعها في إناء ثم خلطها جيداً معاً، ثم اختيار عدد من البطاقات من الإناء مساوٍ لحجم العينة المطلوبة، وتكون العناصر المدونة على البطاقات هي العينة المطلوبة.

ومن الجدير بالذكر أن العينات العشوائية البسيطة يمكن أن تختار بأحد أسلوبين؛ إما السحب بإرجاعٍ، وفيها نقبل التكرار للعناصر المختارة من المجتمع، فقد يظهر نفس العنصر عدد من المرات. وإما السحب بدون إرجاع، وفيها لا نقبل التكرار، فيظهر العنصر مرة واحدة فقط في العينة.

(ب) طريقة العينة الطبقية العشوائية (Stratified Random Sample)

إحدى البدائل لطريقة العينة العشوائية البسيطة، ويمكن استعمالها لزيادة دقة النتائج هي طريقة العينة الطبقية. فإذا شعر الباحث أن نتيجة الدراسة الإحصائية قد تعتمد على العمر أو الجنس أو الدخل أو مستوى التعليم مثلاً. فإنه يقوم بتقسيم المجتمع الإحصائي إلى مجموعات جزئية تعتمد على هذه الصفات وتسمى كل مجموعة طبقة ثم يقوم الباحث باختيار عينة عشوائية بسيطة من كل طبقة، ثم يدمج العينات الجزئية الناتجة لتكون عينة واحدة والتي تسمى بالعينة الطبقية. فإذا أردنا اختيار عينة حجمها 100 من مجتمع مكون من 1000 شخص، ينقسمون إلى ذكور وإناث بنسبة 2 إلى 3 فإنه يجب اختيار من الذكور عينة حجمها 40 ونختار من

الإناث عينة حجمها 60 بطريقة العينة العشوائية البسيطة. ثم ندمج العينتين معاً لنحصل على عينة مكونه من 100 شخص والتي تسمى بالعينة الطبقية.

(ج) العينة العنقودية (Cluster Sample)

من الجدير بالملاحظة أن استعمال أي من الأسلوبين السابقين يتطلب توفير قائمة بعناصر المجتمع الإحصائي. وقد نواجه في كثير من الحالات مواقف لا تتوفر فيها مثل هذه القوائم، لهذا يصبح من المتعذر اختيار عينة من تلك المجتمعات. والبديل لهذا الوضع هو محاولة تشكيل قائمة بالأسماء، وهذا سيؤدى بنا إلى عمل مسح شامل وقد لا يكون بمقدورنا إجراء المسح الشامل. لهذا نلجأ إلى أسلوب آخر وهو تقسيم المجتمع الإحصائي إلى مجموعات جزئية واضحة تسمى كل مجموعة منها عنقوداً، ثم نقوم باختيار عينة عشوائية بسيطة من بين تلك العناقيد. فمثلاً: إذا أردنا دراسة عدد أفراد الأسر في المملكة العربية السعودية، فإننا نقوم بتقسيم المملكة إلى مناطق وكل منطقة يمكن تقسيمها إلى محافظات وكل محافظة يمكن تقسيمها إلى مدن وكل مدينة يمكن تقسيمها إلي أحياء سكنية (أو قرى) أو أية قاعدة أخرى فيصبح لدينا قائمة من المجموعات تسمى كل مجموعة عنقوداً. ثم نقوم باختيار عينة من تلك المجموعات وتجرى الدراسة عليه. ويسمى هذا الأسلوب بأسلوب العينة العنقودية.

(د) طريقة العينة المنتظمة (Systematic Sample)

وهى بديل أخر من بدائل العينة العنقودية وتستعمل في حالة عدم توفر قائمة بعناصر المجتمع الإحصائي. فمثلاً إذا أردنا أن نأخذ عينة من طلاب جامعة الملك سعود الذين يستعملون المطعم الجامعي من أجل التعرف على أسباب استعمال المطعم وعلى مدى جوده خدمات المطعم الجامعي، فإننا سوف نختار تلك العينة من الطلاب الذين يستعملون المطعم عن طريق جلوس شخص على المدخل الرئيسي للمطعم ويقوم بتوجيه بعض الأسئلة، مثلاً لكل عاشر طالب يدخل إلى المطعم فإن العينة التي سنحصل عليها بتلك الطريقة ليست عينة عشوائية بشكل كامل، لأن فيها نوعاً من الانتظام وهو اختيار الطالب العاشر لذا تسمى العينة المختارة بالعينة المنتظمة.

(هـ) طريقة العينة المعيارية (Standard Sample)

عند اختيار العينة يجب توخي الحذر لتكون ممثله للمجتمع الإحصائي الذي أخذت منه والعينة التي تمثل مجتمعها تمثيلاً جيداً هي تلك التي تتفق مقاييسها الإحصائية مع مقاييس المجتمع الإحصائية. فيتفقان في الوسط والوسيط والانحراف المعياري وغيرها. وتسمي العينة التي تتصف بهذه الصفات بعينة معيارية، ويتم اختيار مثل هذه العينات بطريقة تتابعيه وذلك على النحو التالي:

ليكن المطلوب تقدير نسبة نجاح عملية معينة، ففي هذه الحالة نختار عينة من الناس (مرضى أو أصحاء) ونجرى عليهم العملية ثم نقدر نسبة النجاح. ولكن الذي يحدث في الحقيقة أن المرضى يحضرون للمستشفى تباعاً كلما شعر أحدهم بالألم ويقرر المريض بنفسه موافقته على إجراء العملية أو عدم موافقته على إجرائها، وتقوم المستشفى بحفظ سجلات لهؤلاء المرضى ونتيجة العملية الجراحية، وقد تقوم بإيجاد نسبة النجاح لأول 30 شخصاً ثم لأول 60 شخصاً ثم لأول 90 شخصاً أجريت لهم العملية، وهكذا. ويلاحظ الأطباء في النهاية أن نسبة النجاح أصبحت ثابتة تقريباً؛ فعندما تتوقف الدراسة يقومون بنشر نتيجة بحثهم هذا مقدرين نسبة نجاح العملية بالنسبة التي تثبت في النهاية. وتسمى هذه العينة بالعينة المعيارية، لأننا نعتقد بأن لها نفس المقاييس الإحصائية للمجتمع ككل، وذلك بسبب ثبات المقياس بازدياد حجم العينة التدريجي.

البيانات (Data):

تسمى المعلومات التي يتم جمعها وتنظيمها وتحليلها بواسطة الإحصائيين (بيانات (data)) أو (مشاهدات (observations)) ويوجد نوعان من البيانات:

(أ) **البيانات الوصفية (Qualitative Data):** وهى البيانات التي تكون في صورة غير عددية، مثل لون العينين وفصيلة الدم والجنسية.

(ب) **البيانات الكمية (Quantitative Data):** وهى البيانات التي تكون في صورة عددية، مثل الطول والوزن وعدد أفراد الأسرة والدخل ودرجة الحرارة وعمر السلعة وعدد الحوادث الأسبوعية على طريق ما.

ويمكن التمييز بين البيانات سواء كانت لفظيه أو رقميه بناء على ما يسمى بوحدة القياس التي تأخذها البيانات وهى:

١. المقياس الاسمي (Nominal Scale): يستخدم عندما تكون البيانات غير قابله للترتيب، مثل الأكواد الخاصة بالتدخين أو الحالة الاجتماعية... ويعتبر المقياس الأسمى أضعف وحدات القياس.

٢. المقياس الترتيبي (Ordinal Scale): إذا كانت البيانات قابله للترتيب سواء كانت لفظية أو رقمية فإن وحدة القياس الخاصة بها تكون ترتيبية. مثال ذلك ترتيب عدد الأطفال في الأسر، تقديرات مجموعة من الطلاب تصاعدياً أو تنازلياً.

٣. مقياس الفترة (Interval Scale): إذا كانت بيانات المتغير رقمية وكان الفرق بين أي قيمتين من قيم المتغير له معنى (حيث يكون لدينا وحدة قياس ثابتة) فإن نظام القياس يطلق عليه أنه بفترة.

٤. مقياس النسبة (Ratio Scale): عندما تكون الفروض السابقة في مقياس الفترة مستوفاة بالإضافة إلى أن النسبة بين أي قيمتين للمتغير له معنى يكون المقياس بنسبة؛ مثال ذلك مقاييس الوزن والطول والأسعار؛ فالسلعة التي سعرها ثلاثون ريالاً سعودياً سعرها ثلاثة أضعاف السلعة التي سعرها عشرة ريالات سعودية.

المتغير العشوائي (Random Variable):

المتغير العشوائي هو ذلك المتغير الذي تتحدد قيمه بناء على عائد تجربة تخضع نتائجها للصدفة؛ فإذا اخترنا سلعة من إنتاج مصنع فإنها قد تكون جيدة أو تالفة ومن ثم فإن حالة السلعة (X) تأخذ نتيجتين فقط وهي إما $X = 0$ إذا كانت السلعة تالفة أو $X = 1$ إذا كانت السلعة جيدة، ومن ثم فإن حالة السلعة X متغير عشوائي. وإذا أخذنا طالباً من بين طلاب إحدى الكليات فإن طول هذا الطالب (X) يتراوح بين 140سم و 190سم حيث 140سم هي الحد الأدنى لطول الطالب، 190سم هي الحد الأقصى لطول الطالب في الكلية. ويمكن أن يأخذ (X) أي قيمة في الفترة [140, 190] . ومن ثم فإن طول الطالب (X) متغيراً عشوائياً. وإذا اخترنا شخص ما من مطار الملك خالد بالرياض فإنه سيكون له الجنسية (X) السعودية أو المصرية أو السورية أو ... ومن ثم فإن الجنسية (X) متغير عشوائي. ومما سبق فيمكن تقسيم المتغيرات العشوائية إلى نوعين هما المتغيرات العشوائية الكمية والمتغيرات العشوائية الوصفية. يتم تحديد نوع المتغير العشوائي بناء على نوع البيانات التي يأخذها. تنقسم المتغيرات العشوائية الكمية إلى نوعين هما:

(أ) المتغير العشوائي المتقطع (Discrete Random Variable): هو المتغير الذي يأخذ قيماً تقع عند نقاط منفصلة أو هو المتغير الذي تكون قيمه قابلة للعد. مثل عدد أفراد الأسرة $X = 0, 1, 2, ...$ أو عدد الحوادث الأسبوعية على أحد الطرق $X = 0, 1, 2, ...$

(ب) المتغير العشوائي المتصل (Continuous Random Variable): هو المتغير الذي يأخذ أي قيمة في مدى معين، أو هو المتغير الذي لا يمكن حصر جميع قيمه، مثل الطول والعمر والوزن والدخل السنوي لمجموعة من الأشخاص أو درجة الحرارة أو عمر السلعة أو سرعة السيارة بالكيلومتر.

وتتناول الدراسات الإحصائية متغيراً أو أكثر من المتغيرات العشوائية. وتعتمد المتغيرات العشوائية على بعض الثوابت أو المقاييس العددية الوصفية في طبيعتها؛ مثال ذلك متوسط المتغير محسوباً من كل مفردات المجتمع أو نسبة مفردات المجتمع التي تتوفر فيها صفة معينة. وتسمى هذه المقاييس بمعالم المجتمع.

معالم المجتمع (Population Parameters):

معلمة المجتمع هي أي مقياس يصف خاصية من خصائص المجتمع. وهذه المعالم ترافق المتغير العشوائي وهي تميز المجتمعات عن بعضها البعض ونحن نعلم مثلاً أن لكل مجتمع متوسطاً وعلى ذلك فإن متوسط المجتمع معلمة من معالمه ويتحدد المجتمع بمعرفة هذه المعلمة. ولسوء الحظ فإننا لا نقوم عادة بدراسة كل مفردات المجتمع، ولذلك فإن قيم المتغير لجميع مفردات المجتمع تكون مجهولة. ومن ثم تكون القيم الحقيقية لمعالم المجتمع مجهولة. وهنا يجب أن نحصل على تقريب للقيم الحقيقية لهذه المعالم باستخدام البيانات المتوفرة من العينة، ولكي يتم ذلك يجب أن نستخدم ما يسمى بالإحصاءات.

الإحصاءات (Statistics):

الإحصاءة هو مقياس يصف خاصية من خصائص العينة وتتحدد قيمته من مفردات العينة. ومثال هذه الإحصاءات متوسط المتغير محسوب من بيانات العينة (أو اختصاراً: متوسط العينة) وكذلك نسبة مفردات العينة التي تتوفر فيها صفة معينة. وتفيدنا الإحصاءات في أنها مقاييس تصف العينة نفسها وتمكننا من عمل الاستدلال حول معالم المجتمع التي تم اختيار العينة منه.

ونلاحظ أن قيمة الإحصاءات تختلف باختلاف العينة المسحوبة من نفس المجتمع؛ فإذا اخترنا عينة من مجتمع ما وقمنا بحساب المتوسط الحسابي لها ثم اخترنا عينة أخرى من نفس المجتمع فإن المتوسط الحسابي للعينة الثانية قد يختلف عن المتوسط الحسابي للعينة الأولى، وبتكرار عملية اختيار العينات فإننا سنحصل على قيم مختلفة للوسط الحسابي ومن ثم يمكن القول إن الإحصاءات هي متغيرات عشوائية تعتمد على عناصر العينة التي تم اختيارها من المجتمع. ولإجراء أي دراسة إحصائية يجب اتباع عدة خطوات:

١- تحديد الدراسة

٢- تحديد مجتمع الدراسة

٣- تحديد طريقة إجراء الدراسة الإحصائية

٤- عرض البيانات وتبويبها

٥- تحليل بيانات الدراسة

٦- اتخاذ قرار حول الدراسة وإعطاء التوصيات

١-٥ نبذه عن الحزم الإحصائية

تُعدّ الحزم الجاهزة (Package) التي أعدت كتطبيقات في مجالات العلم المختلفة (الطبية، الهندسية، الرياضية، ...) من أهم ما عرف حديثاً كوسيلة لحل أغلب المشاكل، مما ساعد على انتشار استخدام الحاسب الشخصي. وقد ساعدت الحزم الجاهزة على سرعة ودقة وقدرة متخذي القرار مما أدى إلى تطور هائل ومذهل في جميع التخصصات المختلفة وبالأخص في التنبؤ والتخطيط واتخاذ القرار.

وقد مر تطور الإحصاء بعدة مراحل مختلفة من أهمها:

١- ظهور الاحتمالات وهى أحدى اكتشافات الرياضيات التطبيقية الهامة حيث تم التزاوج بين الرياضيات والإحصاء لينتج ما يسمى بالإحصاءات التحليلية التي ساعدت على التنبؤ والتخطيط واتخاذ القرارات.

٢- ظهور الحاسبات التي ساعدت على وضع النظريات الإحصائية موضع التنفيذ الدقيق والسريع.

وبظهور الحاسبات ظهرت مجموعة من الحزم الجاهزة التي تخدم المجالات المختلفة مثل الإحصاء، الهندسة، الطب، الزراعة، الصناعة، الفلك، التجارة، ...، وفي مجال الإحصاء ظهرت مجموعة كبيرة من الحزم الإحصائية منها على سبيل المثال ... MINITAB, MICROSTAT, SATGRAPH, SAS, SPSS, وغيرها.

وسوف نهتم هنا بالتعرف على بعض المفاهيم الأساسية لحزمة SPSS، التي تعد من أكثر الحزم الإحصائية شيوعاً واستخداماً. ومن المعلوم أن الحروف SPSS هي اختصار لـ: Statistical Package for Social Sciences ويبدو أن الحزمة قد أعدت للدراسات الاجتماعية ولكن استخدامها أمتد إلى فروع العلم المختلفة. وقد ظهر على مواقع الانترنت مثل:

http://www.amazon.co.uk/Crash-Course.../dp/1405145315

اختصار أخر لـ SPSS أصبح اختصاراً لـ: Statistical Product and Service Solutions

وظهرت أقدم الإصدارات من الحزمة SPSS عام 1970م، ثم توالت الإصدارات المختلفة لتواكب التقدم العلمي حيث ظهرت الإصدارات 7.0 , 9.0, 10.0, 11.0, 12.0, 13.0, 14.0, 15.0, 16.0, 17.0 وستتوالى الإصدارات لتتابع ما هو جديد في الإحصاء لكي يضاف للحزمة الإحصائية. وسوف نهتم بأن تكون معظم التطبيقات باستخدام برنامج SPSS الذي يعتبر من الأهمية بمكان أن يتعلمه كل إحصائي وأن يتطرق إليه كل دارسي الإحصاء. وتختلف الإصدارات المختلفة لحزمة SPSS بإضافة بعض التطبيقات التي يتم استحداثها وسنقوم بإجراء التطبيقات الإحصائية خلال هذا الكتاب باستخدام الإصدار 17.0 من الحزمة SPSS.

أسئلة وتمارين (١)

١- اشرح مع إعطاء بعض الأمثلة

(أ) المتغير الوصفي

(ب) المتغير الكمي

(ج) المتغير الكمي المتقطع

(د) المتغير الكمي المتصل

٢- أي من المتغيرات التالية وصفية وأي منها كمية:

- لون العينين لمجموعة من الطلاب.
- وزن الطالب في بداية العام الدراسي.
- أسماء محافظات المنطقة الوسطى بالمملكة العربية السعودية.
- درجات الطلاب في الاختبار الفصلي الأول لمقرر 1080 إحص.
- تقديرات الطلاب في المستوى الأول بكلية العلوم والدراسات الإنسانية بالخرج.
- جنسية منسوبي كلية العلوم بالرياض.
- الفترة الزمنية التي يقضيها الطالب بالاختبار النهائي لمقرر 1080 إحص.
- طول المكالمات التليفونية التي يستقبلها شخص ما خلال أسبوع.
- أسماء طلاب المستوى الأول بكلية العلوم والدراسات الإنسانية بالخرج.

٣- حدد المجتمع والمتغير ونوع المتغير وحجم العينة في الفقرات التالية:

- أراد باحث دراسة درجة الحرارة خلال 30 يوماً من أيام السنة.
- في أحد مصانع المصابيح الكهربائية تم حساب زمن حياة 80 مصباحاً من إنتاج المصنع لتقدير عمر المصباح المنتج بواسطة هذا المصنع.
- أراد باحث تحديد نسبة الجنسيات المختلفة الموجدة في المملكة العربية السعودية فقام باختيار 100 شخص من المترددين على إدارة الجوازات بمحافظة ما خلال يومين متتاليين.

- في دراسة لتحديد نوع فصيلة الدم وعلاقتها بمرض ارتفاع ضغط الدم قام باحث باختيار 40 شخصاً من المترددين على المستشفى الجامعي وقام بتعيين نوع فصيلة دمهم.
- في دراسة لاستطلاع الرأي حول مدى موافقة الطلاب على الجدول المقترح للاختبار النهائي في كلية العلوم بجامعة الخرج قامت إدارة الكلية بأخذ رأي 45 طالباً من طلاب الكلية.
- أراد باحث دراسة متوسط درجات الحرارة في فصل الصيف لعام 1431هـ فقام بتسجيل درجات الحرارة خلال 60 يوماً من فصل الصيف.

٤- حدد الإحصاءات والمعالم في العبارات التالية:

- متوسط مجموعة من درجات طلاب كلية العلوم بجامعة الملك سعود.
- نسبة المدخنين بالمنطقة الوسطى بالمملكة العربية السعودية.
- متوسط الحوادث في أحد شوارع مدينة الرياض خلال 30 يوماً في العام.
- متوسط دخل الفرد في المملكة العربية السعودية.
- نسبة المصابين بأنفلونزا الطيور في العالم في عام 1429هـ
- متوسط درجات الحرارة خلال 60 يوماً في صيف 1430هـ.

٥- ما الفرق بين الإحصاءة والمعلمة؟

٦- عرف الإحصاء مع ذكر أنواعه.

٧- ما الفرق بين الإحصاء الحيوي والإحصاء التربوي؟

٨- علم الإحصاء هو علم يخدم معظم العلوم. وضح ذلك.

٩- وضح معنى الاختصار SPSS.

١٠- أي من العبارات التالية صحيح وأيها خطأ:

- يفضل استخدام طريقة المسح الشامل عند إجراء أي دراسة إحصائية.
- المعلمة هي خاصية تميز العينة وهي باستمرار معلومة.
- عدد مرات الإصابة بالأنفلونزا خلال فصل الشتاء هو متغير عشوائي متصل.

١١- ما الفرق بين جمع البيانات باستخدام طريقه المسح الشامل وطريقة العينة؟ وأيهما أفضل؟

١٢- اذكر مميزات وعيوب طريقة المسح الشامل.

١٣- اذكر مميزات وعيوب طريقة العينة.

١٤- ما المقصود بتوزيعات المعاينة؟

١٥- اذكر طرق اختيار العينات.

١٦- اقترح طريقة مناسبة لأخذ عينة فى كل من الحالات التالية:

- اختيار عينة من طلاب كلية العلوم والدراسات الانسانية لتمثيل الكلية فى لقاء مدير جامعة الخرج.
- اختيار عينة من سكان مدينة الرياض لمعرفة مدى رضاهم عن منتج معين.
- اختيار عينة من سكان المملكة العربية السعودية لتقدير نسبة المدخنين فى المملكة.
- اختيار عينة من طلاب الجامعة لدراسة استفتاء حول مدى ضرورة مقرر معين.
- اختيار عينة من مطار الملك خالد بالرياض لتعيين نسبة الوافدين من مصر.

١٧- استعمل جدول الارقام العشوائية لاختيار عينة عشوائية بسيطة حجمها 6 من مجتمع مكون من 45 طالباً.

١٨- ما هو الشرط الضرورى لصحة نظرية النهاية المركزية؟

١٩- ما الفرق بين طريقة العينة العشوائية البسيطة والعينة العنقودية؟

٢٠- متي يتم استخدام طريقة العينة الطبقية؟

٢١- ما الفرق بين العينة المنتظمة والعينة المعيارية؟

الفصل الثاني

تبويب البيانات وتمثيلها بيانياً

DATA TABULATION AND GRAPHICAL REPRESENTATION

٢-١ مقدمة

ذكرنا في الفصل الأول أن البيانات المطلوبة يتم جمعها إما من مفردات المجتمع بالكامل أو من مفردات عينة يتم اختيارها من هذا المجتمع، والبيانات في صورتها الأولية تسمى (بيانات خام). وقد يكون عدد البيانات كبيراً بحيث يصعب الاستفادة منها في حالتها هذه، ففي مثل هذه الحالات يكون من الضروري تلخيص البيانات ووضعها في صورة تسهل عملية الاستفادة منها. وتؤدي عملية تلخيص البيانات التي يتم جمعها إلى التضحية ببعض المعلومات من مفردات المجتمع أو العينة، ولكن ذلك لن يؤدي إلى خسارة كبيرة إلا إذا كنا مهتمين بالمفردات التي تنتمي إليها كل معلومة. ومثال على ذلك: إذا كان اهتمامنا يتركز فقط على الحد الأدنى والحد الأعلى لطول مجموعة من الأشخاص أو على نسبة الأشخاص الذين تتراوح أطوالهم بين 170 و175سم وليس على طول كل شخص في المجموعة.

وسوف نهتم في هذا الفصل بدراسة متغير عشوائي واحد فقط وتلخيص البيانات الخاصة به، وذلك عن طريق:

١- جدولة البيانات الخاصة بالمتغير العشوائي.

٢- التمثيل البياني لقيم المتغير العشوائي.

٣- حساب بعض المقاييس الإحصائية للمتغير العشوائي.

وسوف نتعامل مع كل من المتغيرات العشوائية الوصفية(النوعية) والكمية ووصف بياناتها باستخدام الطرق السابقة بما يتناسب مع طبيعة كل نوع، وسوف تختلف تلك الطرق في كيفية تطبيقها والتي تعتمد في دورها على نوع البيانات.

٢-٢ المتغير العشوائي الوصفي (Qualitative Variable)

لقد تم تعريف المتغير العشوائي الوصفي في الفصل الأول على أنه المتغير الذي يأخذ قيما ليست أرقاماً بل حروفاً أو مجموعة من الحروف(كلمات) مثال ذلك إذا تم تعريف متغير عشوائي X يرمز لنوع فصيلة الدم فإن X سوف يأخذ القيم A , B, O, AB وهي قيم ليست عددية لذا فإن المتغير X هو متغير وصفي كذلك لون البشرة وتقديرات الطلاب ولون الشعر وغير ذلك فهي متغيرات وصفية. توجد عدة طرق من خلالها نستطيع التعرف على خصائص المتغير الوصفي واستخلاص المعلومات منه، وهي إما أن نضع هذه البيانات في جداول تكرارية، أو تمثيلها بيانياً أو حساب بعض المقاييس الإحصائية لها. وسوف نوضح هذه الطرق كل على حده في المثال التالي.

مثال (٢-١)

فيما يلي التقديرات التي حصل عليها 50 طالباً في مقرر 1080 إحص

B	C	A	C	D	F	C	B	A	F
D	C	B	C	F	C	D	C	C	B
C	C	F	C	A	D	B	D	C	C
B	D	C	F	B	C	D	C	B	A
C	C	B	C	D	B	A	F	C	D

ضع هذه البيانات في صورة جدول تكراري ثم مثله بيانياً.

الحل:

نلاحظ أن:

المتغير العشوائي هو: تقديرات الطلاب،

قيم المتغير العشوائي هي: A, B, C, D, F

نوع المتغير: هو متغير عشوائي وصفي(نوعي)

- **الجدول التكراري (Frequency Table)**

يمكن تلخيص البيانات الوصفية وتنظيمها عن طريق حصر لجميع قيم المتغير أو الظاهرة وعدد مرات تكرار كل قيمة من هذه القيم. توضع هذه القيم في جدول يسمى الجدول التكراري لأنه يبين توزيع القيم وتكراراتها. فالمتغير العشوائي X (التقديرات) يأخذ القيم A, B, C, D, F فسوف نضعها في العمود الأول من الجدول. ونضع في العمود الثاني من الجدول العلامات على شكل خط مائل / حيث كل علامة تمثل وجود التقدير المقابل مرة وبتكرار العلامة / عدد من المرات مساوياً لعدد مرات ظهور التقدير، لكننا يجب أن نلاحظ أنه في المرة الخامسة نضع العلامة \ على الأربعة لتكون حزمة وهي عبارة عن مجموعة مكونة من خمس علامات كما يلي:

التقدير (X)	العلامات	التكرار(f)	التكرار النسبي	التكرار المئوي
A	~~////~~	5	5/50 = 0.10	10
B	~~////~~ ~~////~~	10	10/50 = 0.20	20
C	~~////~~ ~~////~~ ~~////~~ ~~////~~	20	20/50 = 0.40	40
D	~~////~~ ////	9	9/50 = 0.18	18
F	~~////~~ /	6	6/50 = 0.12	12
المجموع		50	1.00	100

- **المقاييس الإحصائية (Statistics Measures)**

في حالة البيانات الوصفية المقياس الإحصائي الوحيد الذي يمكن تعيينه لهذه البيانات هو نسبة ظهور أي خاصية أو صفة(p_i) والذي يسمى بالتكرار النسبي للصفة i في حالة البيانات الوصفية.

$$p_i = \frac{f_i}{n}, \quad i = 1,2, \dots, k$$

حيث k تمثل عدد قيم المتغير، f_i عدد مرات ظهور الصفة (التكرار)، n العدد الكلي للبيانات، ومن خواص التكرار النسبي أن $0 \leq p_i \leq 1$ ومجموع العمود الخاص بالتكرار النسبي يساوى الواحد الصحيح. ويمكن أيضا تعيين ما يسمى بالتكرار المئوي c_i للصفة i والذي يمكن حسابه عن طريق ضرب التكرار النسبي في 100

$$c_i = p_i \times 100, i = 1,2, \dots, k$$

والتكرار المئوي يجب أن $0 \leq c_i \leq 100$ ومجموع العمود الخاص بالتكرار المئوي يساوى 100

- **التمثيل البياني (Graphical Representation)**

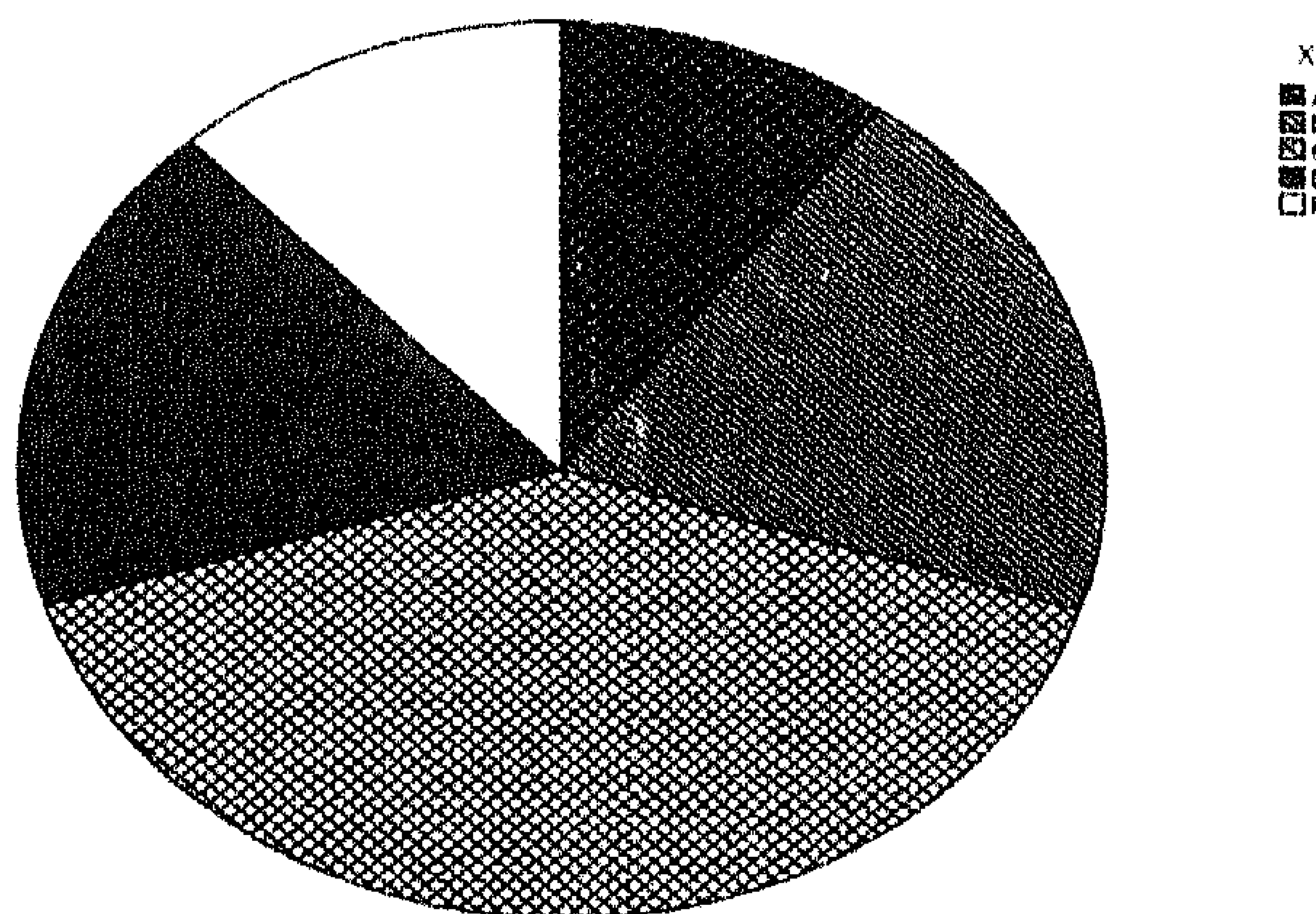
البيانات الوصفية يمكن وصفها بيانياً باستخدام بعض الأشكال البيانية التي تسهل عمليات المقارنة بين قيم المتغير العشوائي ومقارنة قيم نفس المتغير لمجموعتين مختلفتين، والأشكال البيانية التي يمكن استخدامها مع البيانات الوصفية هي الأعمدة البيانية والدائرة البيانية.

(أ) الدائرة البيانية (Pie Charts)

يمكن تمثيل البيانات الوصفية(النوعية) عن طريق الدائرة البيانية، وذلك بتحديد زاوية القطاع لكل صفة (تقدير) ويتم ذلك باستخدام العلاقة التالية:

$$\theta_i = 360 \times p_i = 360 \times \frac{f_i}{n}, i = 1, 2, .., k$$

حيث θ_i تمثل زاوية القطاع المناظرة لكل صفة. وبالرجوع للمثال ورسم القطاعات بالزوايا المناظرة لكل صفة نحصل على الشكل التالي:



شكل (٢-١) الدائرة البيانية لتقديرات الطلاب في مقرر 1080إحص.

حيث:

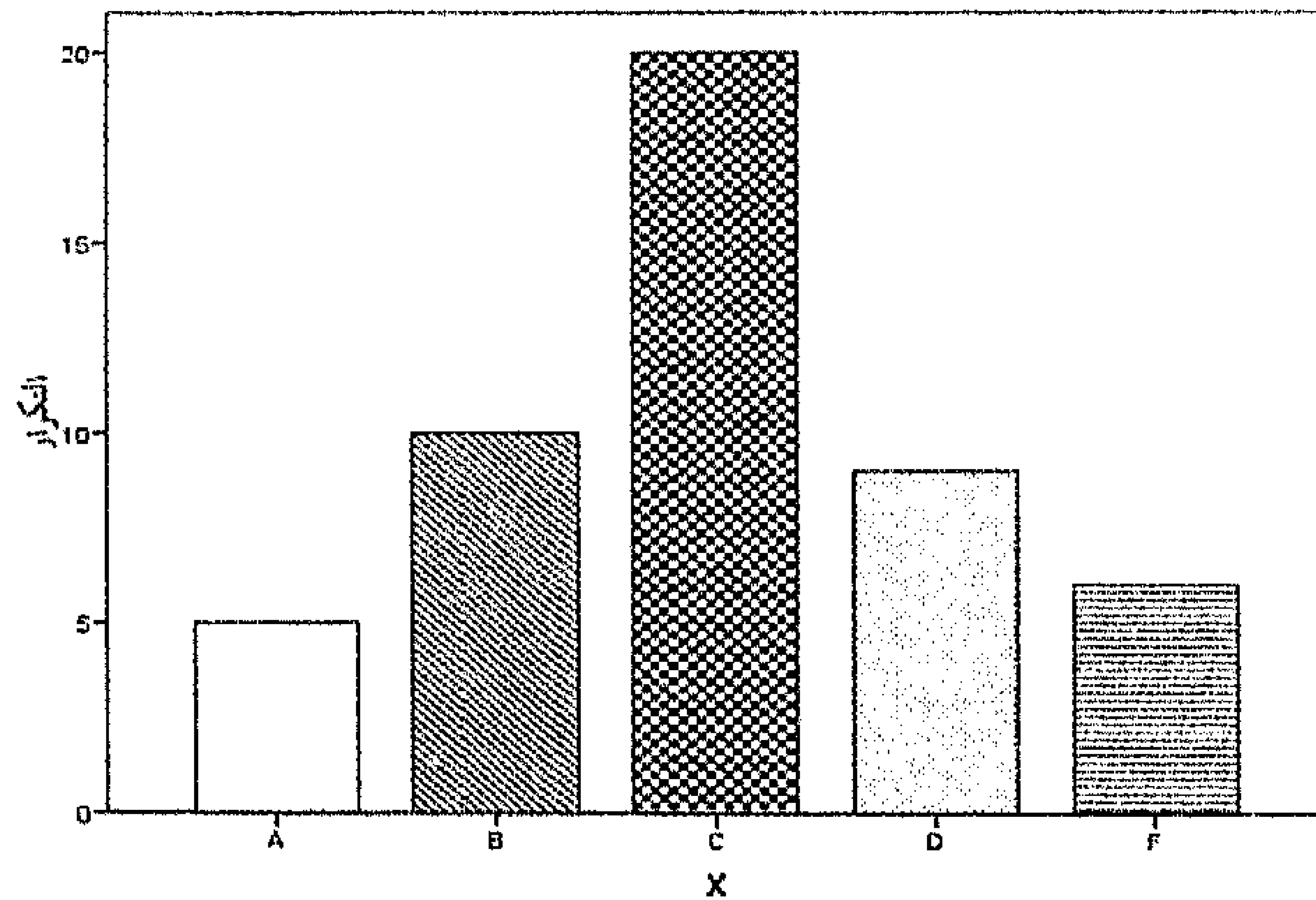
$$\theta_A = 360 \times p_A = 360 \times 0.1 = 36^o$$

$$\theta_B = 360 \times p_B = 360 \times 0.2 = 72^o$$
$$\theta_C = 360 \times p_C = 360 \times 0.4 = 144^o$$
$$\theta_D = 360 \times p_D = 360 \times 0.18 = 64.8^o$$
$$\theta_F = 360 \times p_F = 360 \times 0.12 = 43.2^o$$

من الشكل السابق يمكن مقارنة الصفات (التقديرات) ببعضها البعض عن طريق حجم القطاعات المناظرة للصفات. ونجد أن التقدير C هو أكثر التقديرات تكراراً ثم التقدير B ثم التقدير D ثم التقدير F وأخيراً التقدير A وعملية المقارنة من الرسم البياني أسهل من الجداول التكرارية.

(ب) الأعمدة البيانية (Bar Charts)

الأعمدة البيانية من أكثر الرسومات البيانية استخداماً مع البيانات الوصفية، ويمكن رسم الأعمدة عن طريق رسم المحاور ووضع الصفة على المحور الأفقي والتكرار (أو التكرار النسبي أو التكرار المئوي) على المحور الرأسي لمحاور الإحداثيات ونرسم عمود على كل صفة يكون ارتفاعه هو تكرار تلك الصفة، ويجب أن تكون قواعد الأعمدة متساوية وكذلك أيضا يجب أن تكون المسافات بين الأعمدة متساوية.



شكل(٢-٢). الأعمدة البيانية للتقديرات.

ومن الشكل البياني (شكل ٢-٢) الذي يمثل الأعمدة البيانية لتقديرات الطلاب في المثال (٢-١) يمكن مقارنة التقديرات أي منها أكثر ظهوراً (حدوثاً) من الآخر عن طريق مقارنة الأعمدة الخاصة بتلك التقديرات،

فنلاحظ أن العمود الممثل للتقدير C هو أعلى عمود ثم العمود الممثل للتقدير B، ثم D ثم F وأخيراً العمود الممثل للتقدير A.

ويمكن مقارنة صفة واحدة لأكثر من عينة من البيانات عن طريق الجداول التكرارية أو عن طريق الرسم البياني (الأعمدة)، فعلى سبيل المثال إذا كان لدينا ثلاث شعب لمقرر 1080 إحص وكان الاختبار هو نفس الاختبار (أسئلة الاختبار موحدة) لتلك الشعب فيمكن مقارنة تقديرات الطلاب في الشعب المختلفة عن طريق الجداول التكرارية ومقارنة عدد الطلاب الحاصلين على نفس التقدير ولكنها ستكون عملية مجهدة، لذا توجد طريقة أسهل وهى استخدام الرسم البياني (الأعمدة البيانية) لإجراء تلك المقارنة عن طريق رسم ثلاثة أعمدة لنفس التقدير، كل عمود يمثل عدد الطلاب الحاصلين على هذا التقدير في مجموعة مختلفة ويكون العمود الأعلى هو الأكثر تكراراً. ويجب التأكد من أن حجم العينات متساوٍ، لكن إذا حدث عكس ذلك وكان عدد عناصر العينات(حجم العينات) مختلف فيجب استخدام التكرار النسبي بدلاً من التكرار على المحور الرأسي.

مثال(٢-٢)

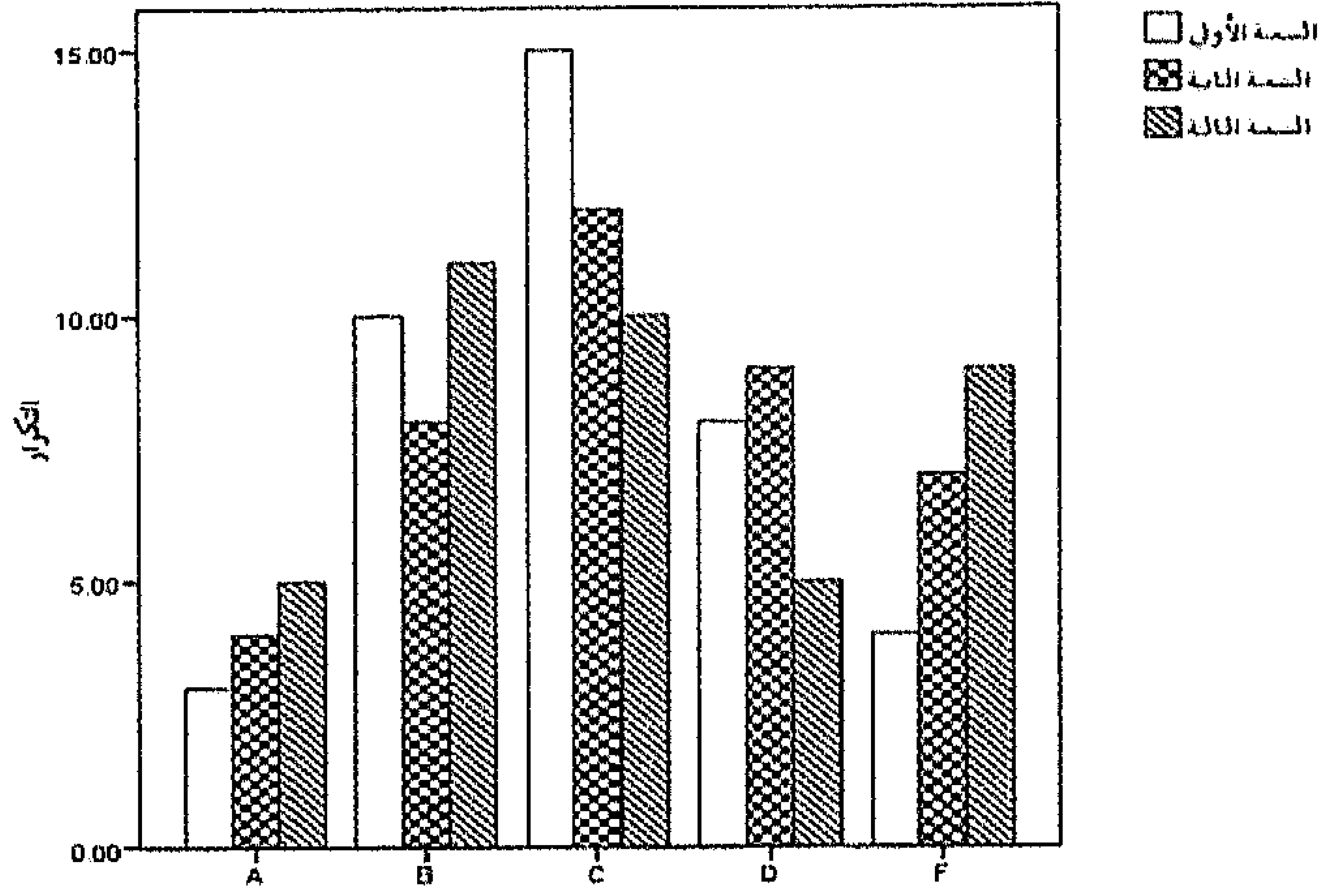
إذا كان لدينا تقديرات للطلاب في ثلاثة شعب لمقرر 1080 إحص علماً بأن كل شعبة بها 40 طالباً وتم تلخيصها في الجدول التكراري التالي:

التقدير	تكرار(الشعبة الأولى)	تكرار(الشعبة الثانية)	تكرار (الشعبة الثالثة)
A	3	4	5
B	10	8	11
C	15	12	10
D	8	9	5
F	4	7	9
المجموع	40	40	40

من الجدول السابق يمكن مقارنة التقديرات المناظرة للشعب المختلفة وذلك عن طريق مقارنة التكرار المناظر لكل صفة في الشعب الثلاثة. ويمكن تمثيل الجدول السابق بيانياً عن طريق الأعمدة البيانية للشعب الثلاث معاً على نفس الشكل.

ومن الشكل البياني (شكل ٢-٣) نلاحظ أن العمود الأول لكل تقدير هو الخاص بالشعبة الأولى والعمود الثاني هو الخاص بالشعبة الثانية والعمود الثالث لكل تقدير هو الخاص بالشعبة الثالثة. ويمكن مقارنة

نفس التقدير للشعب الثلاث بمقارنة الأعمدة الثلاثة الممثلة للتقدير. وأيضا يمكن مقارنة التقديرات المختلفة لشعبة ما بمقارنة العمود الممثل للتقدير بنفس ترتيبه مع التقديرات المختلفة.



شكل(٢-٣) تقديرات الشعب الثلاثة في مقرر 1080إحص

مثال (٢-٣)

تم اختيار عينة من 84 سيارة بطريقة عشوائية من أحد معارض السيارات وتم تلخيص بيانات هذه السيارات حسب أنواعها في الجدول التالي والمطلوب:

(أ) تعيين التكرار النسبي والمئوي لكل نوع.

(ب) تمثيل هذه البيانات بيانياً.

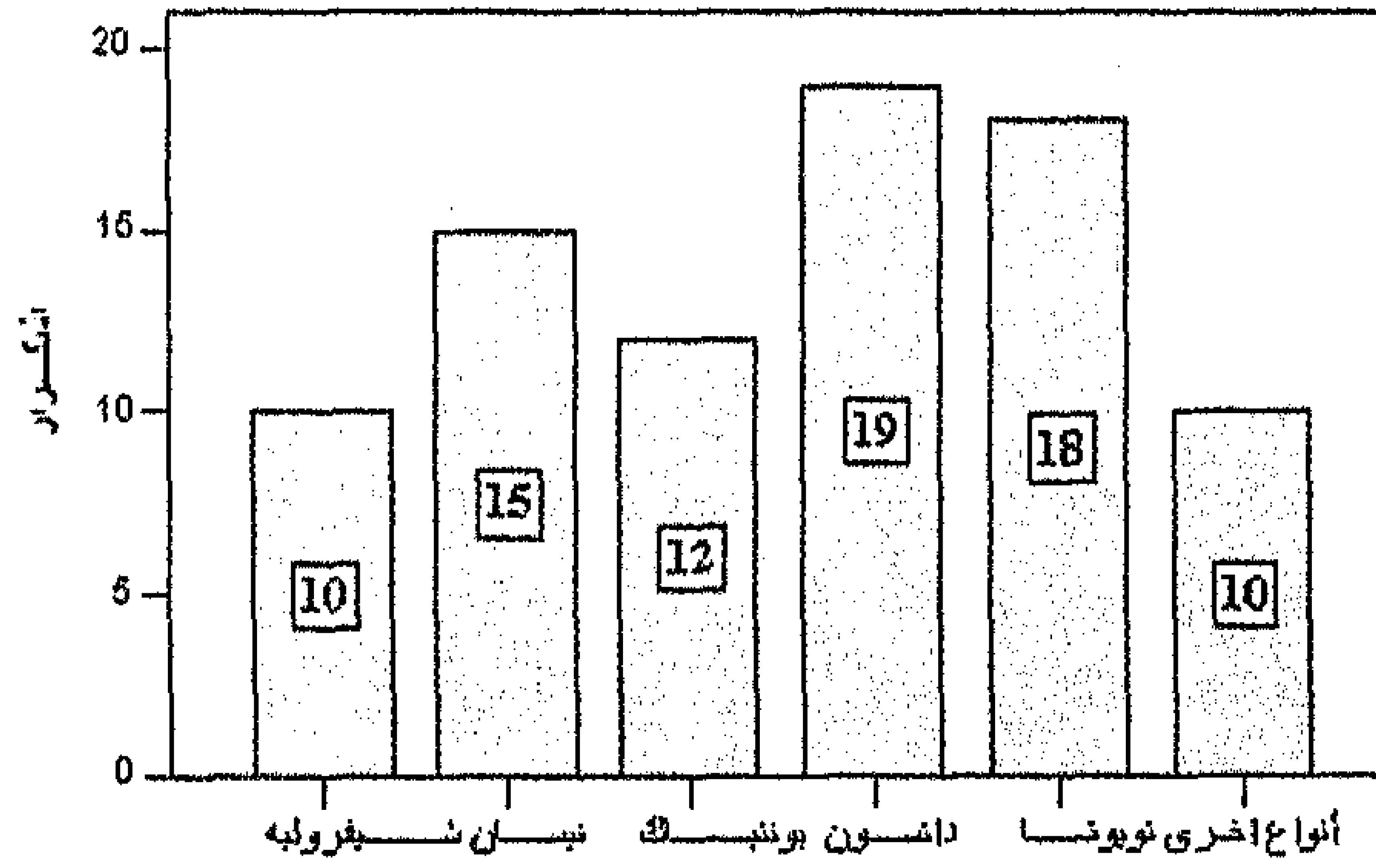
النوع	شيفروليه	نيسان	بونتياك	داتسون	تويوتا	أنواع أخرى
عدد السيارات	10	15	12	19	18	10

الحل:

(أ) التكرار النسبي والمئوي لكل نوع

النوع	التكرار	التكرار النسبي	التكرار المئوي
شيفروليه	10	0.119	11.9%
نيسان	15	0.179	17.9%
بونتياك	12	0.143	14.3%
داتسون	19	0.226	22.6%
تويوتا	18	0.214	21.4%
أنواع أخرى	10	0.119	11.9%
المجموع	**84**	**1.000**	**100%**

(ب) التمثيل البياني لأنواع السيارات



شكل(٢–٤) الأعمدة البيانية لأنواع السيارات.

٢–٣ المتغير العشوائي الكمي (Quantitative Variable)

المتغيرات العشوائية الكمية هي تلك المتغيرات التي تأخذ قيماً عددية، وسوف نهتم هنا بجدوله البيانات والتمثيل البياني لها وسنتحدث عن المقاييس الإحصائية للبيانات الكمية في الفصل التالي.

ولكن قبل التعرض للمتغيرات الكمية يجب أن نفرق بين نوعين من أنواع المتغيرات الكمية فالنوع الأول من المتغيرات الكمية هو ذلك النوع الذي تكون بياناته عبارة عن قيم قابلة للعد ويتم تكرارها مرة على الأقل والذي يسمى بالمتغير الكمي المنفصل، أما النوع الثاني فهو النوع الذي تكون بياناته غير قابله للعد ويسمى بالمتغير الكمي المتصل وسوف نتعرض لكل نوع على حده.

٢-٣-١ المتغير الكمي المنفصل (Discrete Quantitative Variable)

لعرض المتغير الكمي المنفصل باستخدام الجداول التكرارية فإننا سوف نقوم بإنشاء جدول كما حدث مع البيانات النوعية لكن الفرق بينهما أن العمود الأول في هذه الحالة سيحتوى على القيم العددية للمتغير الكمي وسوف نقوم أيضا بتمثيل المتغير الكمي باستخدام الأعمدة البيانية فنقوم بوضع قيم المتغير على المحور الأفقي والتكرار على المحور الرأسي ولكن سيكون العمود الممثل لكل قيمة هو عبارة عن خط مستقيم رأسي. ويمكن توضيح ذلك بالمثال التالي:

مثال (٢-٤)

في عينة من 30 أسرة تم اختيارها من إحدى المدن كان عدد الأطفال في كل أسرة كما يلي:

0	3	0	0	3	0	2	2	0	1
2	1	0	0	1	2	4	0	4	2
1	0	1	0	0	2	0	1	3	2

ضع هذه البيانات في صورة جدول تكراري ومثله بيانياً.

الحل:

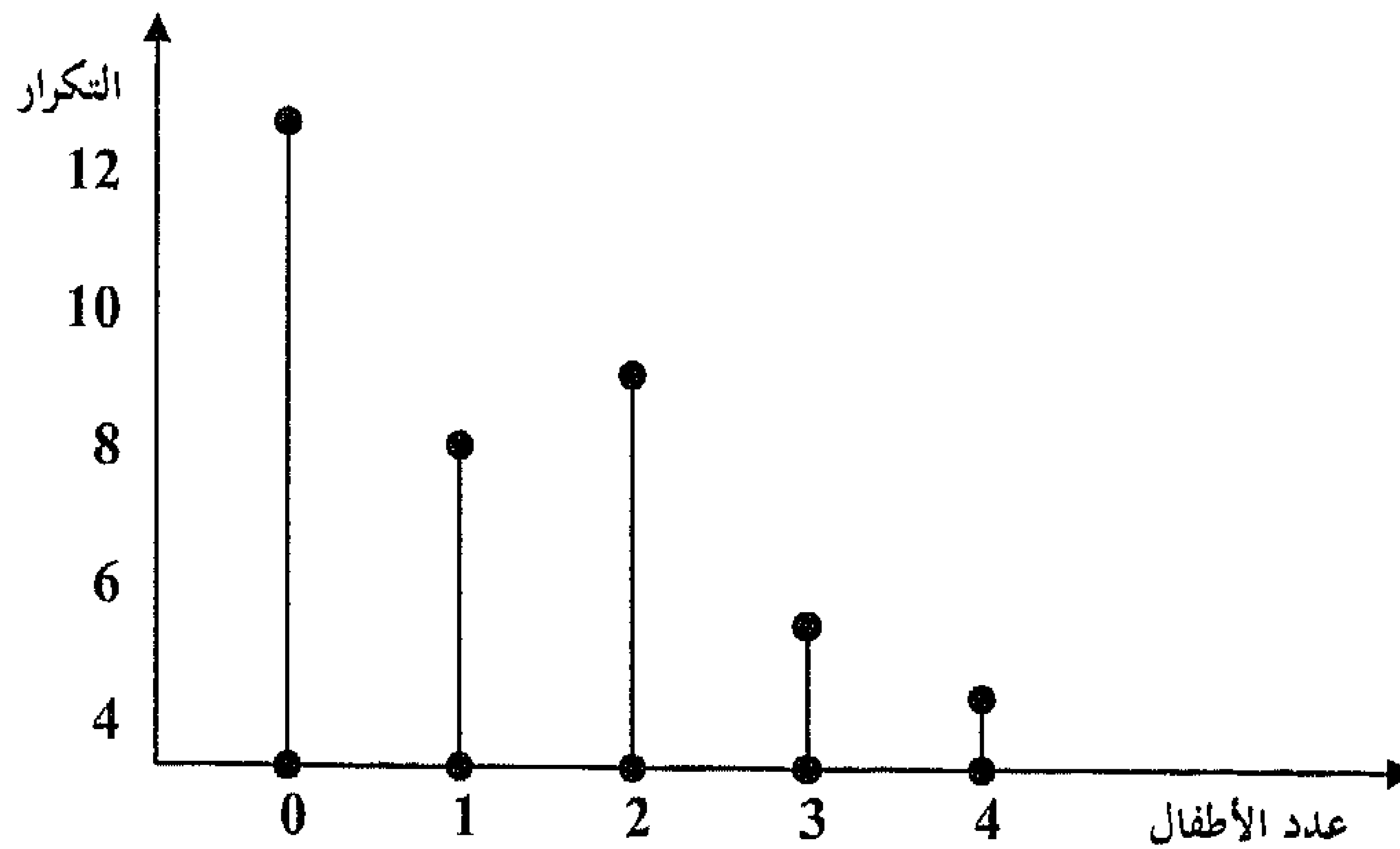
- **الجدول التكراري**

المتغير هو: عدد الأطفال في الأسرة بتلك المدينة، وهنا نجد أن المتغير يأخذ القيم 0, 1, 2, 3, 4 ومن ثم فإن نوع المتغير هو كمي منفصل (متقطع) ويمكن عمل جدول التوزيع التكراري كما يلي:

عدد الأطفال	التكرار	التكرار النسبي
0	12	12/30 = 0.400
1	6	6/30 = 0.200
2	7	7/30 = 0.233
3	3	3/30 = 0.100
4	2	2/30 = 0.067
المجموع	30	1.00

- **التمثيل البياني**

سوف نقوم بتمثيل المتغير الكمي المنفصل عن طريق الأعمدة الرأسية، وسوف تمثل كل قيمة بخط رأسي ارتفاعه هو تكرار القيمة.



شكل (٢-٥). الأعمدة البيانية للمتغير الكمي المتقطع.

٢-٣-٢ المتغير الكمي المتصل (Continuous Quantitative Variable)

(أ) الجدول التكراري (Frequency Table)

المتغير الكمي المتصل هو المتغير الذي تكون قيمة مختلفة، لذا لا يمكن إنشاء جدول يحتوى على القيمة والتكرار فيجب تغيير محتوى العمود الأول عما سبق مع المتغير الكمي المنفصل الذي نقوم فيه بتقسيم مدى المتغير (الفرق بين أكبر وأصغر قيمة) إلى مجموعات متنافية تسمى فئات (classes) أو فترات وسيكون هذا هو محتوى العمود الأول من الجدول التكراري، ولكن العمود الثاني من الجدول سيحتوي على عدد القيم التي تقع داخل كل فترة ويسمى تكرار الفترة، وفي هذه الحالة يسمى الجدول التكراري بـ(الجدول التكراري ذو الفئات أو الفترات). وسوف نتبع الخطوات التالية لإنشاء الجدول التكراري ذو الفترات:

١- تحديد عدد الفترات في الجدول التكراري

عدد الفترات داخل الجدول التكراري يجب أن لا تقل عن 5 فترات ولا تزيد عن 15 فترة ونستخدم الصيغة الآتية في تحديد العدد التقريبي لهذه الفترات(k)

$$k = 1 + 3.322 \times \log_{10}(n)$$

حيث n تمثل عدد مفردات العينة، $\log_{10} n$ هو اللوغاريتم للأساس 10. ويجب أن يكون عدد الفترات عدد صحيحاً، وإذا كان العدد يحتوى على كسر عشري ينبغي أن نقوم بتقريبه للعد الصحيح الأكبر/الأعلى، فمثلاً إذا كان عدد الفترات الناتج من القانون هو 6.02 فنقوم بتقريبه إلى العدد 7

٢- حساب المدى للبيانات وهو الفرق بين أكبر قيمة X_{max} وأصغر قيمة X_{min} فيكون المدى(R) مساوياً:

$$R = X_{Max} - X_{Min}$$

٣- تحديد طول الفترة (w) وهو ناتج قسمة المدى على عدد الفئات

$$w = \frac{R}{k}$$

ويجب أن يكون عدد الأرقام العشرية في طول الفئة مساوياً لعدد الأرقام العشرية في البيانات نفسها، وعلى ذلك يتم تقريب الحد الأدنى لطول الفئة.

٤- تحديد وحدة الدقة للبيانات u وهو أقل رقم يحتوى على نفس العدد من الخانات مساوياً لعدد خانات البيانات.

- فإذا كانت البيانات أعداداً صحيحة فأقل رقم صحيح هو الواحد لذلك وحدة الدقة هي $u = 1$.
- إذا كانت البيانات تحتوى على رقم واحد فقط بعد العلامة العشرية (الفاصلة) فإن أصغر رقم يحتوى على رقم واحد بعد العلامة هو 0.1 وهنا تكون وحدة الدقة هي $u = 0.1$.
- إذا كانت البيانات تحتوى على رقمين فقط بعد العلامة العشرية(الفاصلة) فإن أصغر رقم يحتوى على رقمين بعد العلامة هو 0.01 وهنا تكون وحدة الدقة هي $u = 0.01$، وهكذا.

٥- تحديد الحد الأدنى التقريبي للفئة الأولى وسيكون أقل رقم في البيانات إذا لم يحدد مسبقاً

$$L_1 = X_{Min}$$

٦- تحديد الحد الأعلى التقريبي للفئة الأولى وسيكون عبارة عن الحد الأدنى مضافاً إليه طول الفترة بعد طرح وحدة الدقة منه

$$U_1 = L_1 + w - u$$

٧- تعيين حدود الفترات التالية، وذلك عن طريق إضافة طول الفترة على حدود الفترة السابقة لتعيين الفترة التالية وبتكرار ذلك حتى نحصل على k من الفترات

$$L_i = L_{i-1} + w, \quad U_i = U_{i-1} + w, i = 2, 3, 4, \ldots, k$$

٨- تحويل الفترات التقريبية لفترات فعلية(وتسمي أحياناً بالفترات الحقيقية) وذلك عن طريق طرح نصف وحدة الدقة من الحدود الدنيا للفترات وإضافة نصف وحدة الدقة للحدود العليا للفترات

$$L_i^* = L_{i-1} - \frac{u}{2},\ \ U_i^* = U_{i-1} + \frac{u}{2}\ , i = 1, 2, \dots, k$$

٩- تفريغ البيانات بالجدول وتعني تعيين عدد القيم التي توجد في كل فترة والذي يسمى بتكرار الفترة.

مثال (٢-٥)

بفرض أن لدينا مجموعة من البيانات عددها 50 قيمة وكانت هذه البيانات كلها أعداداً صحيحة بحيث كانت أصغر قيمة هي 11 وأكبر قيمة هي 37 والمطلوب تكوين الفترات التقريبية والفعلية لهذه البيانات.

الحل: من البيانات المعطاة نجد أن

$$n = 50, X_{Min} = 11, X_{Max} = 37$$

ومن الملاحظ أن البيانات كلها أعداد صحيحه ومن ثم سيكون أقل عدد صحيح هو 1، ولهذا فوحدة الدقة هي $u = 1$، عدد الفترات في الجدول التكراري هي:

$$k = 1 + 3.322 \log_{10}(\mathrm{n}) = 1 + 3.322 \log_{10}(50) = 6.644 \cong 7$$

المدى للبيانات هو:

$$R = X_{Max} - X_{Min} = 37 - 11 = 26$$

طول الفترة:

$$w = \frac{R}{k} = \frac{26}{7} = 3.71429 \cong 4$$

بداية أول فترة تقريبية هو:

$$L_1 = 11$$

نهاية أول فترة تقريبية هو:

$$U_1 = L_1 + w - u = 11 + 4 - 1 = 14$$

ومن ثم يمكن تعيين حدود الفترة الثانية بإضافة طول الفترة w على حدود الفترة الأولى لتكون

$$L_2 = L_1 + w = 11 + 4 = 15$$
$$U_2 = U_1 + w = 14 + 4 = 18$$

وهكذا يمكن تعيين حدود الفترات التالية بنفس الطريقة حيث

$$L_i = L_{i-1} + w\ \ , U_i = U_{i-1} + w\ ,\ \ \ i = 2,3, \dots, 7$$

وسوف نحصل على الفترات التقريبية التالية

الفترات التقريبية
11 – 14
15 – 18
19 – 22
23 – 26
27 – 30
31 – 34
35 – 38

ويمكن تعيين حدود الفترات الفعلية باستخدام العلاقة

$$L_i^* = L_i - \frac{u}{2} = L_i - \frac{1}{2}, \quad i = 1,2, \dots, 7$$
$$U_i^* = U_i + \frac{u}{2} = U_i + \frac{1}{2}\ , \quad i = 1,2, \dots,7$$

وسوف نحصل على الفترات التالية:

الفترات الفعلية(أو الحقيقية)	الفترات التقريبية
10.5 – 14.5	11 – 14
14.5 – 18.5	15 – 18
18.5 – 22.5	19 – 22
22.5 – 26.5	23 – 26
26.5 – 30.5	27 – 30
30.5 – 34.5	31 – 34
34.5 – 38.5	35 – 38

ونلاحظ أنه في حالة الفترات التقريبية يوجد فاصل بين نهاية أي فترة وبداية الفترة التالية لها وتلك القيمة هي وحدة الدقة، لكن في حالة الفترات الفعلية(الحقيقية) لا يوجد فاصل بين نهاية وبداية فترتين متتاليتين، ومن ذلك المنطلق يمكننا التعرف على الفترات هل هي فترات تقريبية أم فترات فعلية وتحديد وحدة الدقة إذا كانت الفترات تقريبية حتى يمكن تحويلها لفترات فعلية(حقيقية). وفى بعض الأحيان تكتب الفترات الفعلية بدون كتابة الحدود العليا للفئات كالتالي:

الفترات الفعلية
10.5 –
14.5 –
18.5 –
22.5 –
26.5 –
30.5 –
34.5 – 38.5

وذلك لأن نهاية الفترات الفعلية هي بداية الفترة التالية لها. وهذه الصورة لا تصلح إلا في حالة الفترات الفعلية فقط، وعليه فإذا كتبت الفترات بتلك الصورة فيجب أن نعلم أن هذه الفترات فعلية(حقيقية) وليست تقريبية.

مثال (٢-٦)

البيانات التالية تمثل أوزان 57 طفل بالرطل في أحد مراكز الرعاية الصحية بالمملكة العربية السعودية

68	63	42	27	30	23	36	28	32	79	46	27
22	23	24	25	44	19	65	43	25	74	30	51
36	42	28	31	28	43	25	45	12	57	12	51
12	32	49	38	42	49	27	31	50	38	28	21
16	24	69	47	23	49	22	43	27			

بفرض أن الحد الأدنى لأول فترة هو 10 والمطلوب تكوين الجدول التكراري لأوزان هؤلاء الأطفال مبيناً التكرار النسبي والمئوي لكل فترة.

الحل:

المتغير هو: وزن الأطفال ويأخذ المتغير قيماً عددية ومن ثم فإن المتغير هو متغير كمي متصل

حجم العينة 57 قيمة للوزن وبالتالي فإن عدد الفترات في الجدول التكراري

$$k = 1 + 3.322 \times \log_{10}(n) = 1 + 3.322 \times \log_{10}(57) = 6.833 \cong 7$$

مدى البيانات:

$$R = X_{Max} - X_{Min} = 79 - 12 = 67$$

طول الفترة:

$$w = \frac{R}{k} = \frac{67}{7} = 9.57 \cong 10$$

وحدة الدقة: $u = 1$ ومن ثم فإن الحدود التقريبية والفعلية للفترات هي

الحدود التقريبية	الحدود الفعلية
10 – 19	9.5 – 19.5
20 – 29	19.5 – 29.5
30 – 39	29.5 – 39.5
40 – 49	39.5 – 49.5
50 – 59	49.5 – 59.5
60 – 69	59.5 – 69.5
70 – 79	69.5 – 79.5

تفريغ البيانات بمعنى تحديد عدد القيم بكل فترة(التكرار)

الحدود التقريبية	الحدود الفعلية	التكرار	التكرار النسبي	التكرار المئوي
10 – 19	9.5 – 19.5	5	0.0877	8.77%
20 – 29	19.5 – 29.5	19	0.3333	33.33%
30 – 39	29.5 – 39.5	10	0.1754	17.54%
40 – 49	39.5 – 49.5	13	0.2281	22.81%
50 – 59	49.5 – 59.5	4	0.0702	7.02 %
60 – 69	59.5 – 69.5	4	0.0702	7.02%
70 – 79	69.5 – 79.5	2	0.0351	3.51%
المجموع		57	1.00	100%

مثال (٢–٧)

لدراسة كمية الحديد التي تستهلكها الإناث البالغات اللاتي لا يزيد عمرهن عن 45 سنة، تم اختيار عينة منهن عددها 45 أنثى فكانت كمية الحديد (بالمليجرام) التي تتناولها كل منهن خلال فترة طولها 24 ساعة هي:

15.0	18.1	14.4	14.6	10.9	18.1	18.2	18.3	15.0
16.0	12.6	16.6	20.7	19.8	11.6	12.8	15.6	11.0
15.3	9.4	19.5	18.3	14.5	16.6	11.5	16.4	12.5
14.6	11.9	12.5	18.6	13.1	12.1	10.7	17.3	12.4
17.0	6.3	16.8	12.5	16.3	14.7	12.7	16.3	11.5

ضع هذه البيانات في صورة جدول تكراري.

الحل

المتغير هو: كمية الحديد التي تناولتها الإناث البالغات خلال 24 ساعة،

قيم المتغير هي مجموعة من الأعداد ومن ثم فإن المتغير هو متغير كمي متصل،

حجم العينة 45 قيمة ومن ثم فإن عدد الفترات في الجدول التكراري هي

$$k = 1 + 3.322 \times \log_{10}(n) = 1 + 3.322 \times \log_{10}(45) = 6.492 \cong 7$$

مدى البيانات:

$$R = X_{Max} - X_{Min} = 20.7 - 6.3 = 14.4$$

وحدة الدقة: (من الملاحظ أن البيانات تحتوى على رقم واحد فقط بعد العلامة العشرية) ومن ثم فإن

$$u = 0.1$$

طول الفترة:

$$w = \frac{R}{k} = \frac{14.4}{7} = 2.05714 \approx 2.1$$

الحد الأدنى التقريبي للفترة الأولى:

$$L_1 = X_{Min} = 6.3$$

الحد الأعلى التقريبي للفترة الأولى:

$$U_1 = L_1 + w - u = 6.3 + 2.1 - 0.1 = 8.3$$

ومن ثم الحد الأدنى والأعلى الفعلي للفترة الأولى هو

$$L_1^* = L_1 - \frac{u}{2} = 6.3 - \frac{0.1}{2} = 6.25$$

$$U_1^* = U_1 + \frac{u}{2} = 8.3 + \frac{0.1}{2} = 8.35$$

ومن ثم فإن حدود الفترات التقريبية والفعلية هي:

الفترات الفعلية(الحقيقية)	الفترات التقريبية
6.25 – 8.35	6.3 – 8.3
8.35 – 10.45	8.4 – 10.4
10.45 – 12.55	10.5 – 12.5
12.55 – 14.65	12.6 – 14.6
14.65 – 16.75	14.7 – 16.7
16.75 – 18.85	16.8 – 18.8
18.85 – 20.95	18.9 – 20.9

ومن البيانات يمكن تحديد عدد القيم التي توجد بكل فترة وسنحصل على جدول التوزيع التكراري كما يلي:

التكرار النسبي	التكرار	الفترات الفعلية(الحقيقية)	الفترات التقريبية
0.022	1	6.25 – 8.35	6.3 – 8.3
0.022	1	8.35 – 10.45	8.4 – 10.4
0.267	12	10.45 – 12.55	10.5 – 12.5
0.178	8	12.55 – 14.65	12.6 – 14.6
0.244	11	14.65 – 16.75	14.7 – 16.7
0.200	9	16.75 – 18.85	16.8 – 18.8
0.067	3	18.85 – 20.95	18.9 – 20.9
1.000	45	المجموع	

(ب) التمثيل البياني

إلى جانب تبويب البيانات الكمية المتصلة فإن هناك طرق أخرى تستخدم لتلخيص هذه البيانات، ألا وهي عرض هذه البيانات بشكل بياني. فقد يكون هذا الشكل البياني أكثر تعبيراً من كتابة هذه الصفحات من البيانات. وتوجد عدة طرق لتمثيل البيانات الكمية المتصلة منها:

المدرج التكراري (Frequency Histogram)

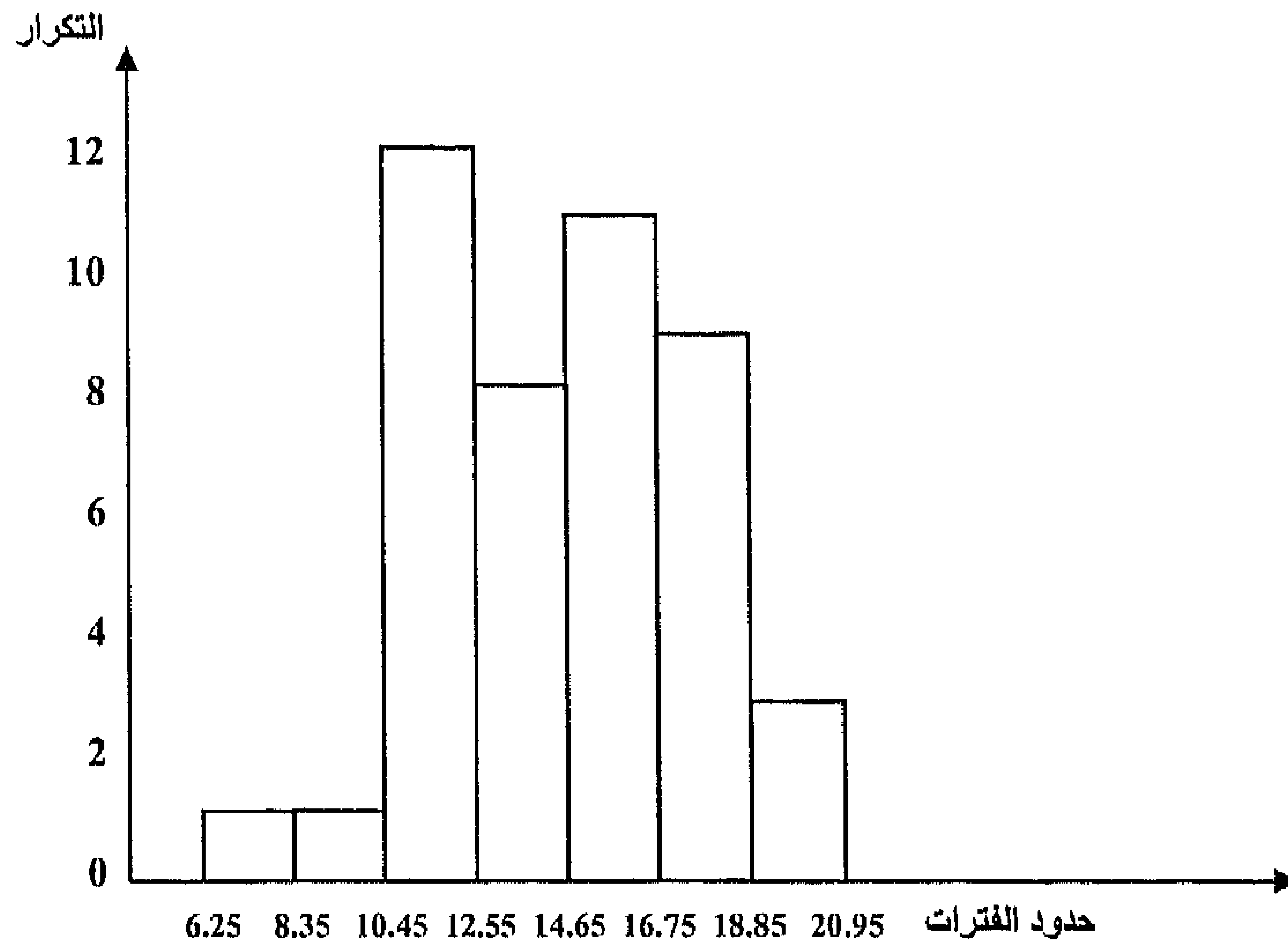
إذا كانت البيانات التي لدينا تمثل متغيراً كمياً متصلاً والتي تقسم عادة إلى فئات (classes)، فإن الشكل البياني الذي يناسبها هو المدرج التكراري. وهو يشبه الأعمدة الراسية، ولكن مساحة الأعمدة في المدرج التكراري مهمة جداً لأنها تتناسب مع التكرار. ولرسم المدرج التكراري فإننا نقسم المحور الأفقي إلى أجزاء مناظرة للفترات الفعلية ثم نرسم لكل فترة عموداً قاعدته هي طول الفترة وارتفاعه هو تكرار هذه الفترة.

مثال (٢-٨)

ارسم المدرج التكراري لمثال (٢-٧).

الحل:

نقوم برسم المحاور ونضع الحدود الفعلية للفترات على المحور الأفقي والتكرار على المحور الرأسي ونقوم برسم عمود قاعدته هي الفترة وارتفاعه هو التكرار المقابل لتلك الفترة.



شكل (٢-٦). الأعمدة البيانية للمتغير الكمى المتصل.

ويمكن أيضا رسم المدرج للتكرارات النسبية (relative frequency histogram) باستخدام التكرارات النسبية بدلاً من التكرارات، وكذلك رسم المدرج التكراري للتكرارات المئوية أيضاً.

وفي بعض الحالات تكون أطوال فئات(فترات) التوزيع التكراري غير متساوية ويحدث ذلك عندما نلاحظ أن أغلب المشاهدات تتركز في جزء صغير من المدى. في مثل هذه الحالات يجب رسم المدرج التكراري باستخدام التكرارات المعدلة وذلك حتى تكون مساحات الأعمدة متناسبة.

$$\text{التكرار المعدل للفئة} = \frac{\text{التكرار الحقيقى للفئة}}{\text{طول الفئة}}$$

مثال (٢-٨)

الجدول الآتي يبين توزيع 80 سلعة حسب أعمارها بالأسابيع والمطلوب رسم المدرج التكراري لها.

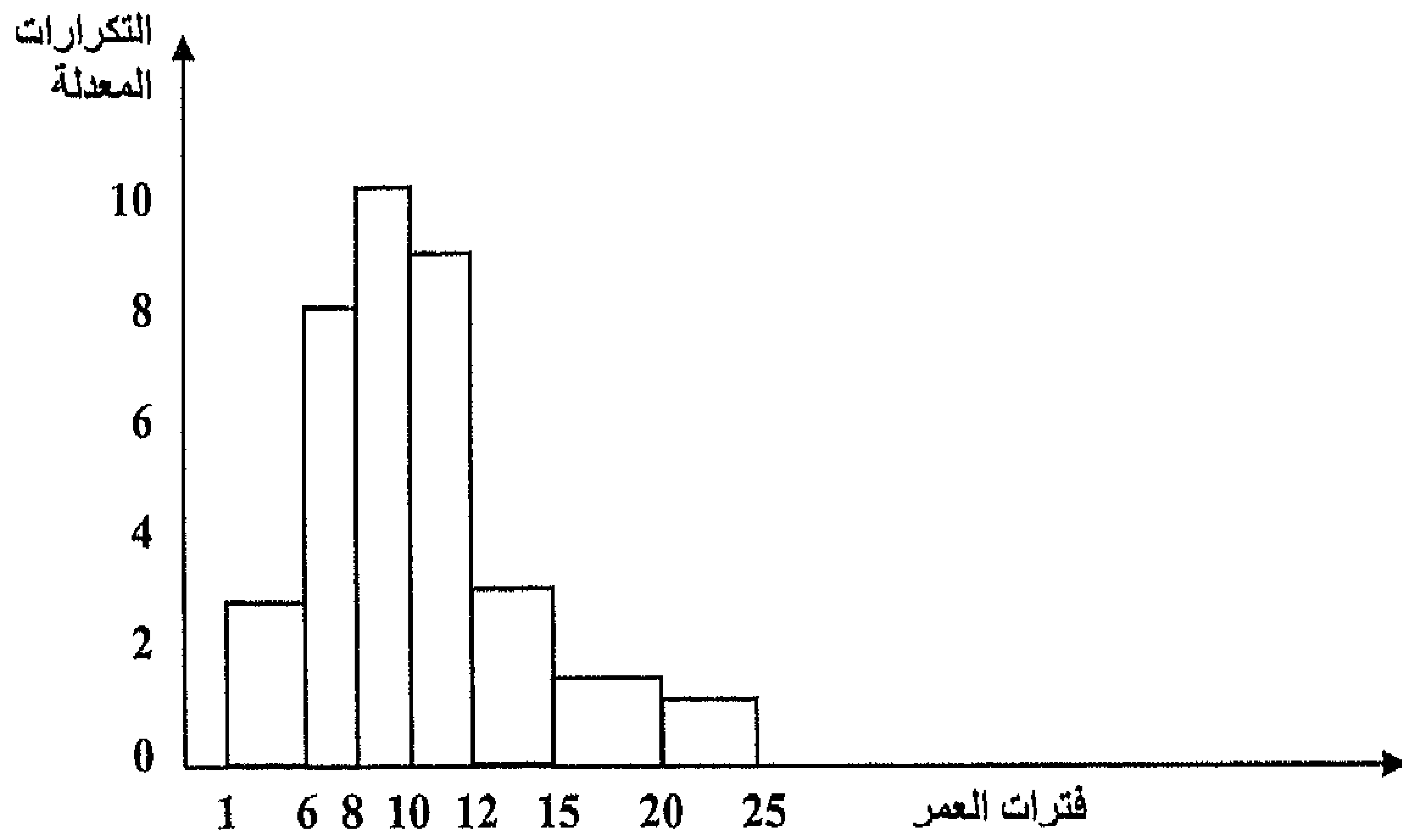
فترات العمر	1 –	6 –	8 –	10 –	12 –	15 –	20 – 25
عدد السلع	14	16	20	18	9	7	5

الحل

نلاحظ أن فترات التوزيع التكراري في هذا المثال أطوالها غير متساوية ومن ثم يجب الحصول على التكرارات المعدلة لاستخدامها في رسم المدرج التكراري.

فترات العمر	عدد السلع	أطوال الفترات	التكرارات المعدلة
1 –	14	5	2.8
6 –	16	2	8.0
8 –	20	2	10.0
10 –	18	2	9.0
12 –	9	3	3.0
15 –	7	5	1.4
20 – 25	5	5	1.0

وبرسم المدرج التكراري باستخدام التكرارات المعدلة مع الفترات نحصل على الشكل البياني التالي:



شكل (٢–٧). المدرج التكراري لتوزيع السلع.

تعريف(٢–١):

مركز الفترة هي القيمة التي تتوسط الفترة ونحصل عليها بقسمة مجموع طرفي الفترة على 2 بمعنى

$$x_i = \frac{L_i + U_i}{2}, \quad i = 1, 2, \dots, k$$

ونلاحظ أن مركز الفترات لا يعتمد على نوع الفترات سواء فترات تقريبية أو فترات فعلية.

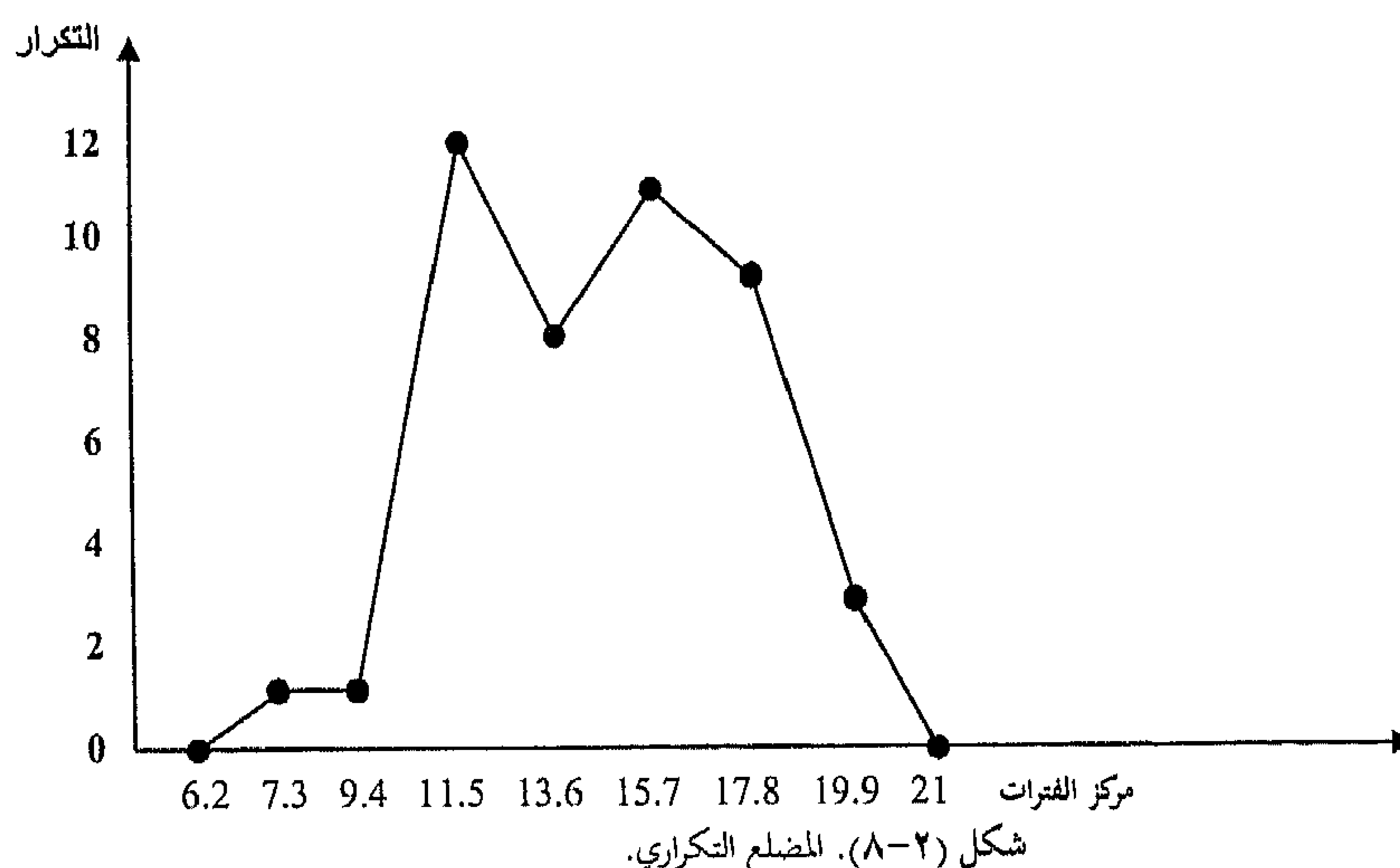
وبالعودة للبيانات الموجودة في مثال (٢–٧) يمكن حساب مراكز الفترات كما هو موضح بالجدول التالي:

مركز الفترات	التكرار	الفترات الفعلية	الفترات التقريبية
7.3	1	6.25 -8.35	6.3 – 8.3
9.4	1	8.35 - 10.45	8.4 – 10.4
11.5	12	10.45 - 12.55	10.5 – 12.5
13.6	8	12.55 - 14.65	12.6 – 14.6
15.7	11	14.65 - 16.75	14.7 – 16.7
17.8	9	16.75 - 18.85	16.8 – 18.8
19.9	3	18.85 - 20.95	18.9 – 20.9
	45	المجموع	

المضلع التكراري (Frequency Polygon)

يُعد المضلع التكراري من الطرق المستخدمة في عرض البيانات بيانياً، ولرسم المضلع التكراري فإننا نقوم برسم محورين أحداهما أفقي والآخر رأسي كما هو الحال في المدرج التكراري. وتمثل البيانات بنقاط كل نقطة إحداثيتها هما مركز الفترة وتكرار الفترة ثم نضيف فترة في البداية وأيضا فترة في النهاية كل منهما تكرارها صفر. ثم نقوم بتوصيل هذه النقاط بخطوط مستقيمة فنحصل على المضلع التكراري, إذا تم التوصيل بين النقاط بمنحنى وليس بخط مستقيم فيسمى الشكل الناتج بالمنحنى التكراري وليس المضلع التكراري.

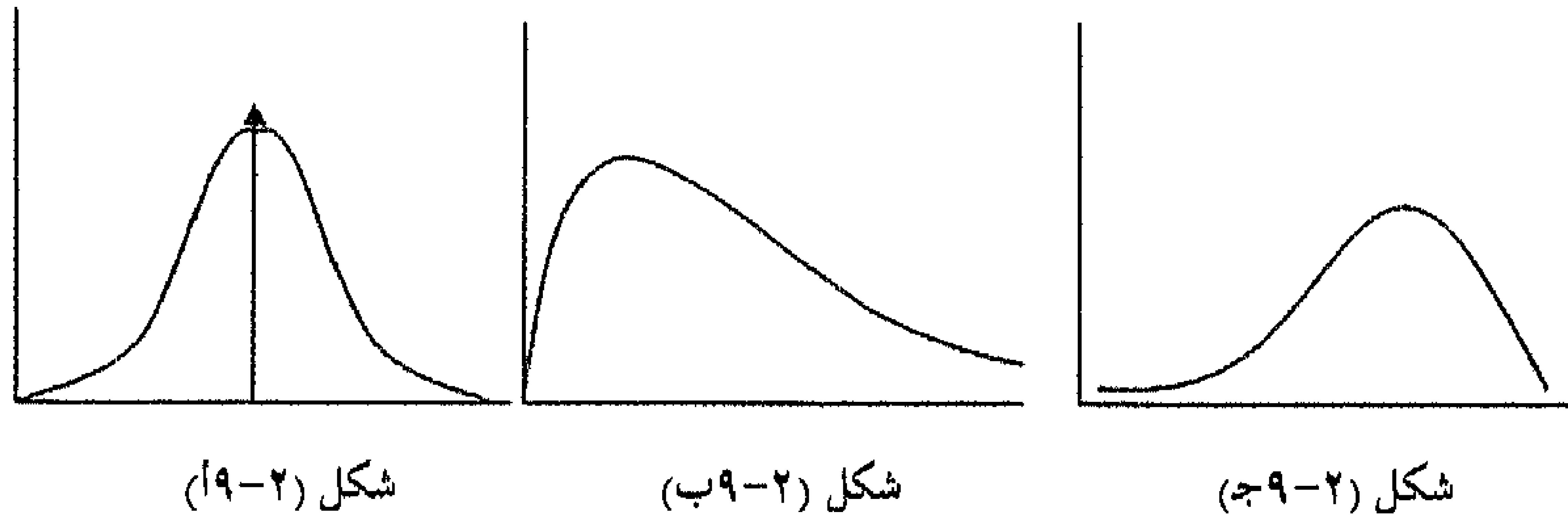
ويمكن رسم المضلع التكراري لمثال (٢-٧) كما هو موضح بالشكل(٢-٨).



شكل (٢-٨). المضلع التكراري.

ويمكن رسم كل من المدرج التكراري والمضلع التكراري في شكل بياني واحد لنجد أنه في حالة الفترات المتساوية يتساوى مجموع مساحات الأعمدة في المدرج التكراري مع المساحة الموجودة تحت المضلع التكراري. لكن إذا استخدمت التكرارات النسبية في حالة المضلع التكراري فإن المساحة تحت المضلع تساوى الواحد الصحيح، ويقترب شكل المدرج التكراري أكثر فأكثر من شكل المنحنى، وفى النهاية يتحول المدرج التكراري إلى منحنى تكراري.

وتأخذ المنحنيات التكرارية أشكالاً مختلفة، فمنها ما يكون له قمة واحدة قرب منتصف مدى المتغير ويتناقص المنحنى من الجانبين إما بالمعدل نفسه وإما بمعدلين مختلفين في اتجاه بداية ونهاية المتغير. فإذا كانت قمة المنحنى تقع عند منتصف مدى المتغير ويتناقص المنحنى من الجانبين بمعدل متماثل فإن المنحنى يكون متماثلا (symmetric curve) شكل (٢-٩أ) .



شكل (٢-٩أ) شكل (٢-٩ب) شكل (٢-٩جـ)

ونسبة كبيرة من التوزيعات التكرارية يكون لها مثل هذا المنحنى. وإذا كان معدل التناقص في المنحنى أسرع جهة اليسار منه جهة اليمين بحيث يكون الجانب الأيمن من المنحنى أطول من الجانب الأيسر شكل(٢-٩ب) فان المنحنى يكون ملتويا ناحية اليمين (positively skewed curve) وإذا كان معدل التناقص في المنحنى أسرع جهة اليمين منه جهة اليسار بحيث يكون الجانب الأيسر من المنحنى أطول من الجانب الأيمن شكل (٢-٩جـ) فإن المنحنى يكون ملتويا جهة اليسار (negative skewed curve).

٢-٤ التوزيعات التكرارية المتجمعة (Cumulative Frequency Distributions)

الجدول التكراري يعطى التكرار المقابل لكل فترة، الممثل بعدد القيم التي تقع في تلك الفترة، لكننا قد نرغب أحياناً في معرفة عدد المشاهدات التي تقل عن كل قيمة من قيم المتغير أو تلك التي تساوي أو تزيد عن كل قيمة من قيم المتغير وهذه الأعداد تسمى التكرارات المتجمعة. فإذا قمنا بحساب عدد المشاهدات التي تقل عن

كل قيمة من قيم المتغير فإننا نحصل على التكرارات المتجمعة الصاعدة، لكن إذا حسبنا عدد المشاهدات التي تساوي أو تزيد عن كل قيمة من قيم المتغير فإننا نحصل على التكرارات المتجمعة النازلة (الهابطة). ويمكن أيضا حساب التكرارات النسبية المتجمعة أو التكرارات المتجمعة النسبية بقسمة كل تكرار متجمع على عدد المفردات. وسوف نعرض كيفية إنشاء كل من التكرار الصاعد وأيضاً الهابط في المثال التالي.

مثال (٢-١٠)

أوجد التوزيع التكراري المتجمع الصاعد والتوزيع التكراري المتجمع الهابط للجدول التكراري الآتي:

طول النبات بالسنتيمتر	20 –	25 –	30 –	35 –	40 –	45 – 50
عدد النباتات	12	16	30	25	12	5

الحل

التوزيع التكراري المتجمع الصاعد:

لإنشاء التكرار المتجمع الصاعد نتبع الخطوات التالية:

١- يجب أن يكون لدينا فترات فعلية وليست تقريبية.

٢- نكون جدولاً من عمودين، العمود الأول نضع فيه الحدود العليا للفترات وتكتب (أقل من الحد العلوي) والعمود الثاني نضع فيه مجموع التكرارات ابتدأ من بداية البيانات حتى الحد العلوي للفترة والجدول التالي يوضح ذلك

الحدود العليا للفترات	التكرار المتجمع الصاعد
أقل من 20	0
أقل من 25	12
أقل من 30	28
أقل من 35	58
أقل من 40	83
أقل من 45	95
أقل من 50	100

التوزيع التكراري المتجمع الهابط:

لإنشاء التكرار المتجمع النازل (الهابط) نتبع الخطوات التالية:

١- نتأكد من أن الفترات فعلية وليست تقريبية

٢- نكون جدولاً من عمودين، العمود الأول نضع فيه الحدود الدنيا للفترات وتكتب (الحد الأدنى فأكثر) نضع في العمود الثاني مجموع التكرارات ابتدأ من بداية الحد الأدنى للفئة حتى أخر البيانات. والجدول التالي يوضح ذلك

التكرار المتجمع النازل	الحدود الدنيا للفترات
100	20 فأكثر
88	25 فأكثر
72	30 فأكثر
42	35 فأكثر
17	40 فأكثر
5	45 فأكثر
0	50 فأكثر

٢-٥ التمثيل البياني للتوزيعات التكرارية المتجمعة

يمكن عرض التوزيعات التكرارية المتجمعة بيانياً باستخدام المضلعات التكرارية المتجمعة، وذلك عن طريق رسم محورين أحدهما أفقي والآخر رأسي، نضع على المحور الأفقي حدود الفترات الفعلية ونضع نقطة مع نهاية الفترة ارتفاعها يمثل التكرار المتجمع الصاعد(الهابط) ثم نقوم بتوصيل تلك النقاط معاً، ففي حالة التوزيع التكراري الصاعد نحصل على منحنى متجمع صاعد ويكون مغلقاً من البداية ومفتوحاً من النهاية أما في حالة التوزيع التكراري المتجمع الهابط فنحصل على منحنى متجمع هابط يكون مفتوحاً من البداية ومغلقاً من نهايته. ويستخدم هذان المنحنيان في الحصول على بعض القيم (التكرارات) التي يكون من الصعب الحصول عليها من التوزيع التكراري العادي. ويتضح ذلك من المثال التالي.

مثال (٢-١١): في المثال السابق (٢-١٠)

(أ) ارسم المنحنى المتجمع الصاعد ومنه أوجد عدد النباتات التي تقل أطوالها عن 37سنتيمتراً.

(ب) إذا علمت أن النبات يكون غير صالح إذا قل طوله عن حد معين. فإذا كانت 10% من النباتات غير صالحة فما الحد الأدنى لطول النبات الصالح؟

(ج) ارسم المنحنى المتجمع الهابط ومنه أوجد عدد النباتات التي يبلغ طولها 32 سنتيمتراً أو أكثر

(د) إذا علمت أن 15% من النبات يزيد طوله عن حد معين فما هذا الحد؟

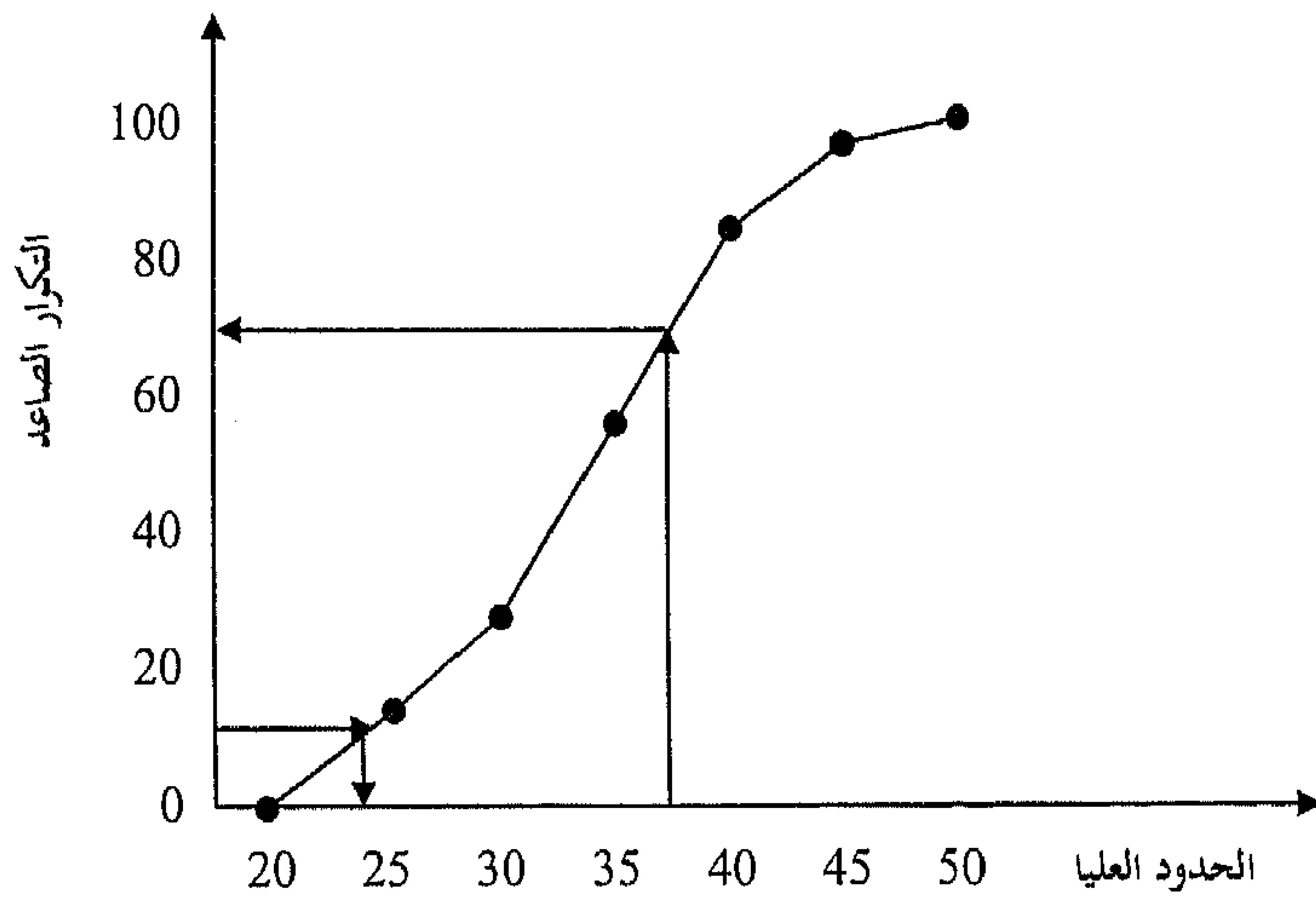
(هـ) أوجد نسبة النباتات التي يتراوح طولها بين 27 سنتيمتراً ،37 سنتيمتراً

الحل

(أ) رسم المنحنى المتجمع الصاعد شكل (٢-١٠)

من شكل (٢-١٠) المحور الأفقي الذي يمثل أطوال النباتات (الفترات) نقوم برسم خط مستقيم رأسي من الطول 37 يقابل منحني التكرار المتجمع الصاعد في نقطة منها نقوم بإسقاط خط أفقي يقابل المحور الرأسي (التكرار الصاعد) في نقطه تمثل العدد المطلوب. ومن ثم فإن عدد النباتات التي يقل طولها عن 37 سنتيمتراً (من الرسم) = 73 نباتاً

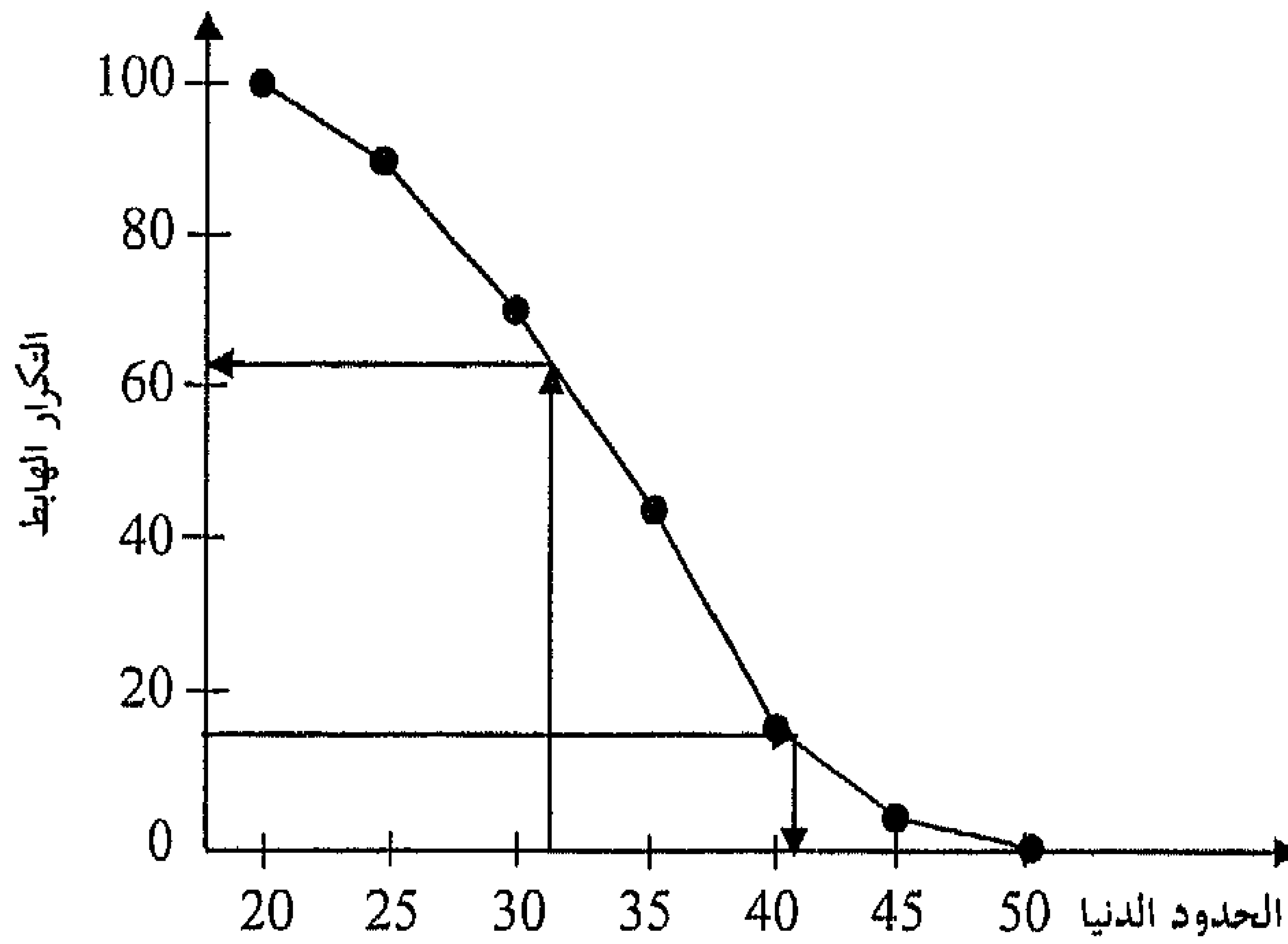
(ب) عدد النباتات غير الصالحة= $\frac{10}{100}\times 100 = 10$ نبتات



شكل (٢-١٠). المنحنى المتجمع الصاعد

من المحور الرأسي الذي يمثل عدد النباتات (التكرار المتجمع) نقوم بإسقاط خط مستقيم أفقي من العدد 10 يقابل المنحني الصاعد في نقطة منها نقوم بإسقاط خط رأسي يقابل المحور الأفقي (حدود الفترات) في نقطه تمثل طول النبات المطلوب. ومن ثم فإن الحد الأدنى لطول النبات الصالح (من الرسم) = 24 سنتيمترا

(ج) رسم المنحنى المتجمع الهابط، شكل (٢-١١)



شكل (٢-١١). المنحنى المتجمع الهابط.

من المحور الأفقي الذي يمثل أطوال النباتات (الفترات) نقوم بإسقاط خط مستقيم رأسي من الطول 32 يقابل المنحني التكرار في نقطه منها نقوم بإسقاط خط أفقي يقابل المحور الرأسي (التكرار النازل) في نقطة تمثل العدد المطلوب. ومن ثم فإن عدد النباتات التي يبلغ طولها 32 سنتيمترا أو أكثر (من الرسم) = 63 نبتة

(د) 15% من النباتات = $\frac{15}{100} \times 100 = 15$ نبتة

من المحور الرأسي الذي يمثل عدد النباتات (التكرار الهابط) نقوم بإسقاط خط مستقيم أفقي من الرقم العدد 15 يقابل المنحني في نقطة منها نقوم بإسقاط خط رأسي يقابل المحور الأفقي (الفترات) في نقطة تمثل طول النبات المطلوب. ومن ثم فإن الحد الأدنى لطول هذه النباتات (من الرسم) = 41 نبتة

أي أن 15% من النباتات يبلغ طولها 41 سنتمتراً أو أكثر.

(هـ) يمكن الحصول على عدد النباتات التي يتراوح طولها بين 27 سنتيمترا و 37 سنتيمترا باستخدام المنحنى المتجمع الصاعد أو المنحنى المتجمع الهابط فمن (أ) نجد أنه باستخدام التكرار المتجمع الصاعد فإن عدد النبتات التي

يقل طولها عن 37 سم هي 73 نبته وبنفس الطريقة يمكن تعيين عدد النبتات التي يقل طولها عن 27 سم وهي 20 نبتة ومن ثم فإن عدد النبتات التي يتراوح طولها بين 27 سم، 37سم يساوى

$$73 - 20 = 53$$

ونسبتها المئوية هي

$$\frac{53}{100} \times 100 = 53\%$$

٢-٦ تطبيقات باستخدام برنامج SPSS

في هذا الجزء سوف نتعرض لكيفية وصف المتغيرات الوصفية والكمية باستخدام برنامج SPSS أي إننا سوف نقوم بعرض كيفية:

١- إنشاء الجداول التكرارية

٢- التمثيل البياني

وسنقوم بحل التمارين التي تم عرضها سابقاً باستخدام برنامج العرض والتحليل الإحصائي SPSS:

أولاً المتغير العشوائي الوصفي (Qualitative Variable)

تطبيق (٢-١) :

في مثال (٢-١) والذي يمثل التقديرات التي حصل عليها 50 طالباً في مقرر 1080 إحص

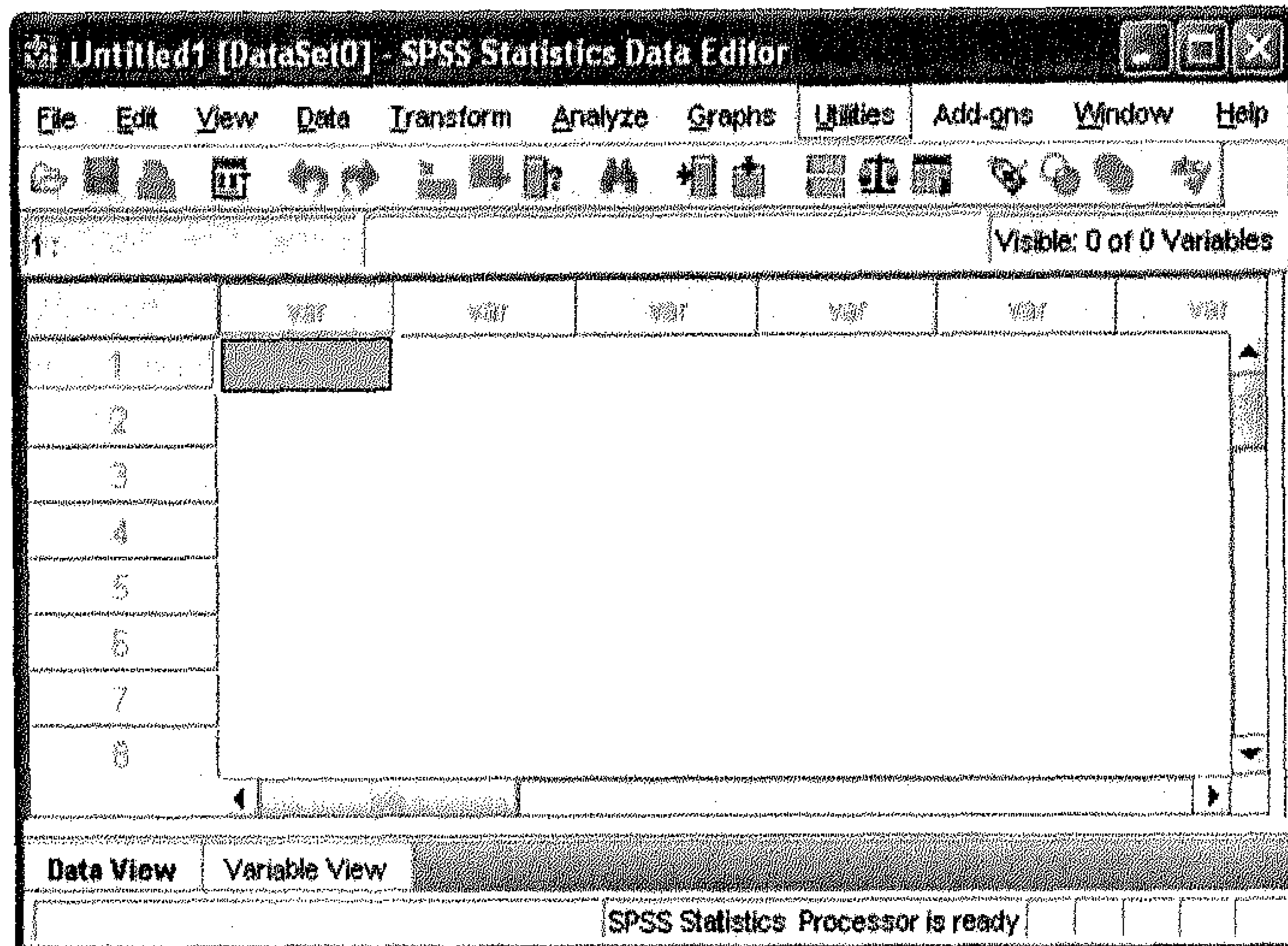
B	C	A	C	D	F	C	B	A	F
D	C	B	C	F	C	D	C	C	B
C	C	F	C	A	D	B	D	C	C
B	D	C	F	B	C	D	C	B	A
C	C	B	C	D	B	A	F	C	D

ضع هذه البيانات في صورة جدول تكراري ثم مثله بيانياً

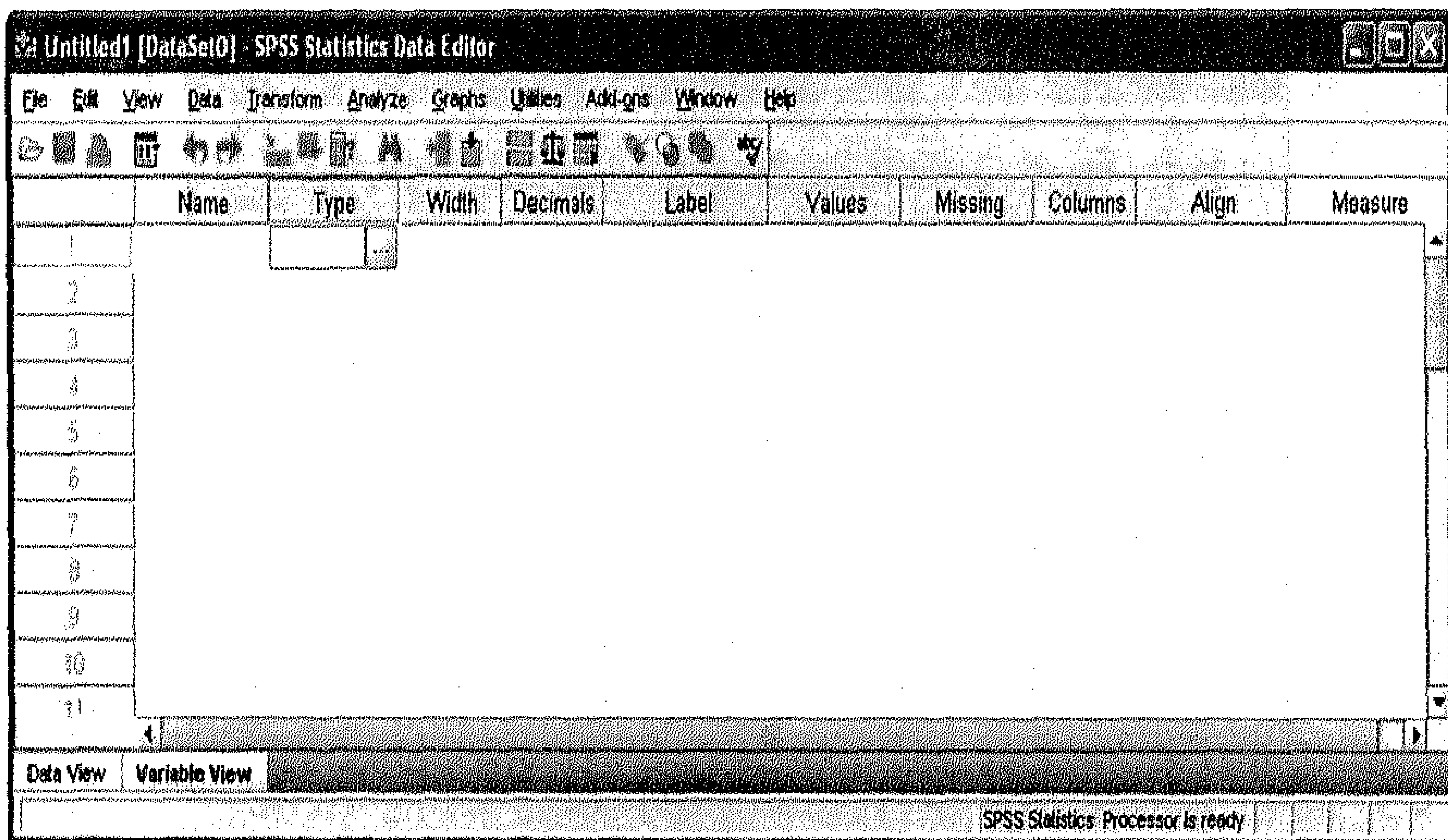
الحل:

سوف نتبع الخطوات التالية لحل هذا التمرين

١- بعد تشغيل برنامج SPSS نقوم بفتح ملف جديد



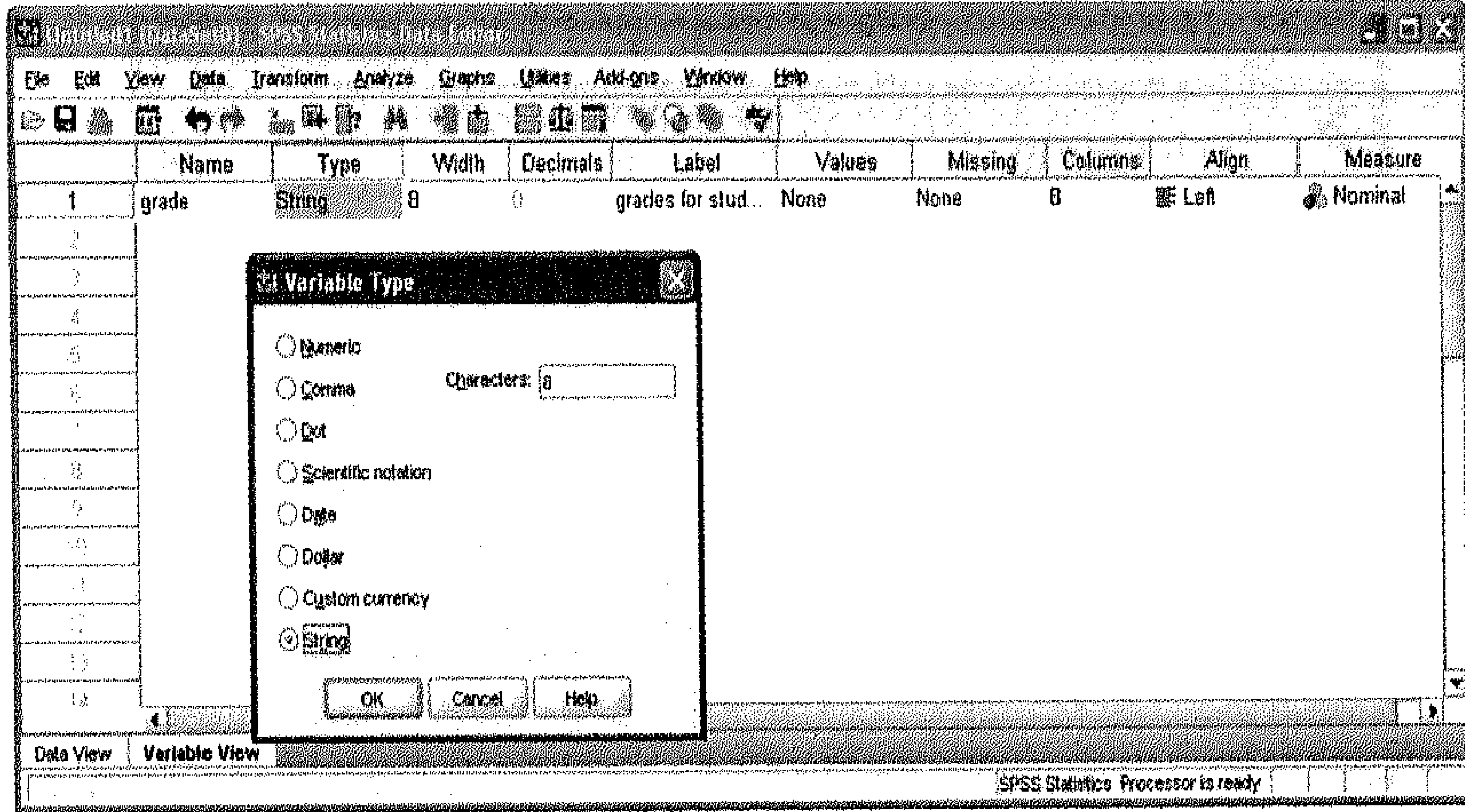
٢- يجب أولا تعريف متغير التقدير وذلك بالضغط على Variable View الموجود في الركن الأيسر أسفل الشاشة وستظهر الشاشة التالية:



٣- نحدد كلا من

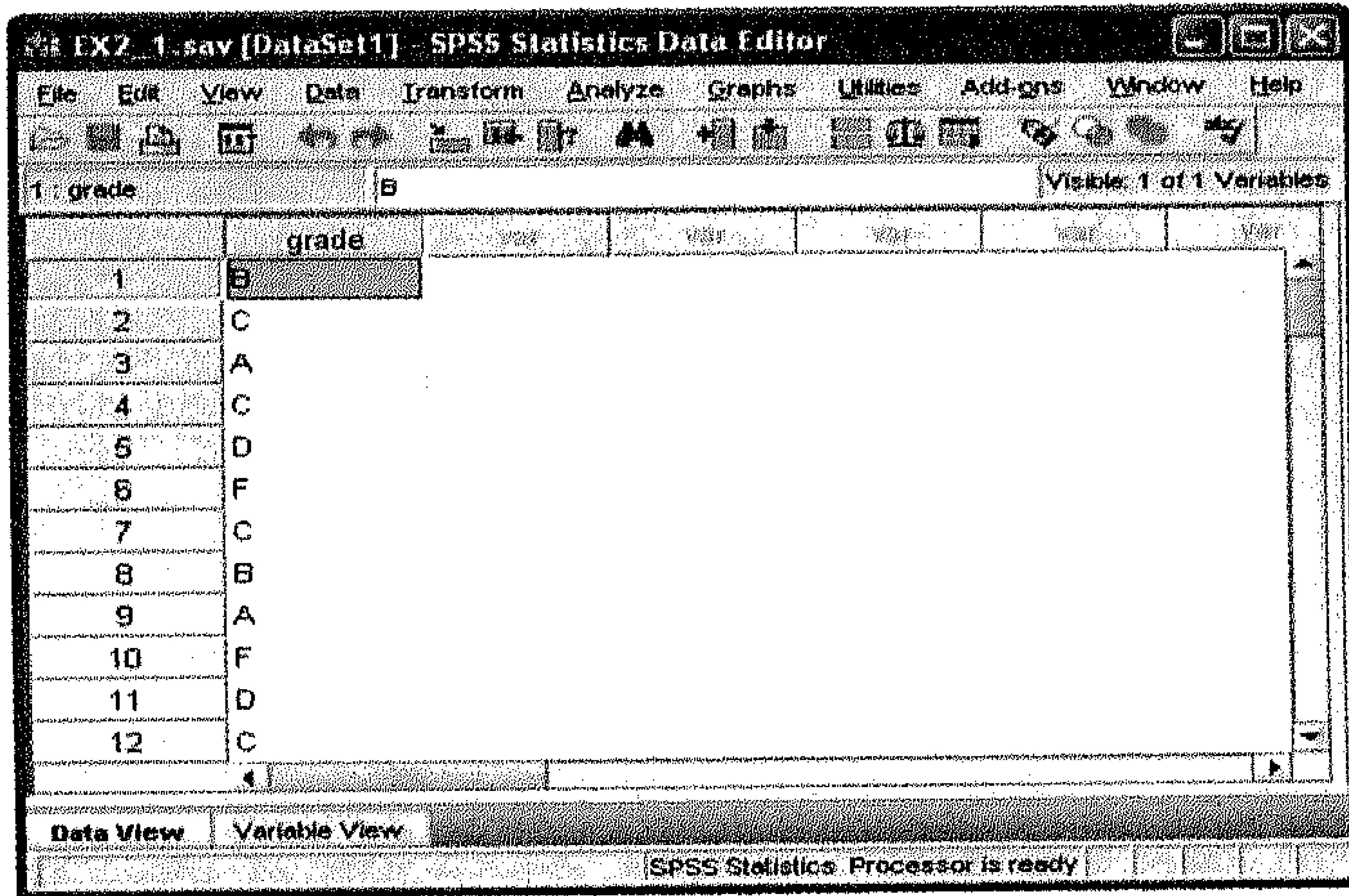
- اسم المتغير وليكن Grade

- نوع المتغير وهو متغير وصفى string
- ويمكن إضافة Label للمتغير وليكن Grades for students in stat 1080

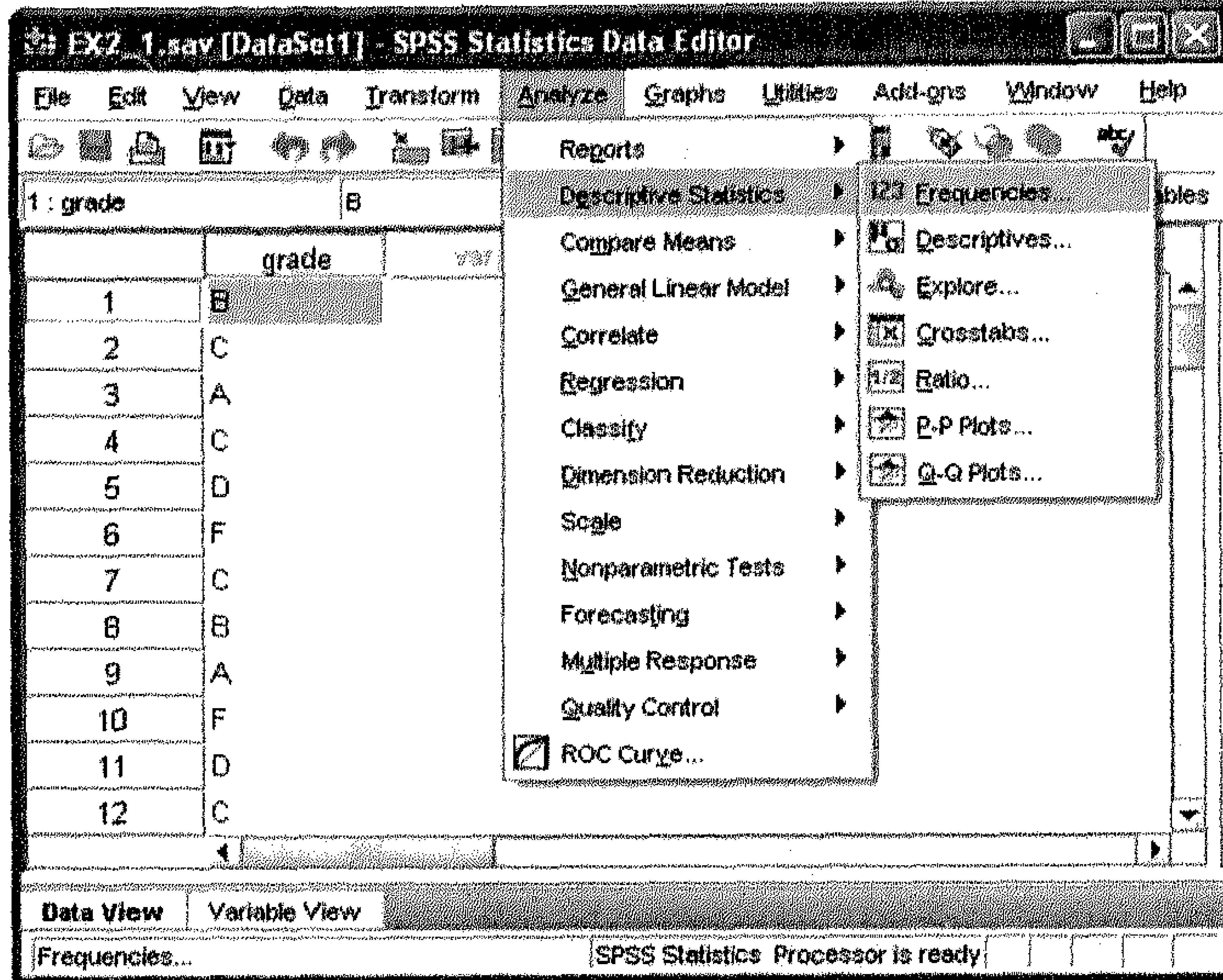


٤- نتنقل لنافذة Data View بالضغط على الاختيار Data View أسفل يسار الشاشة

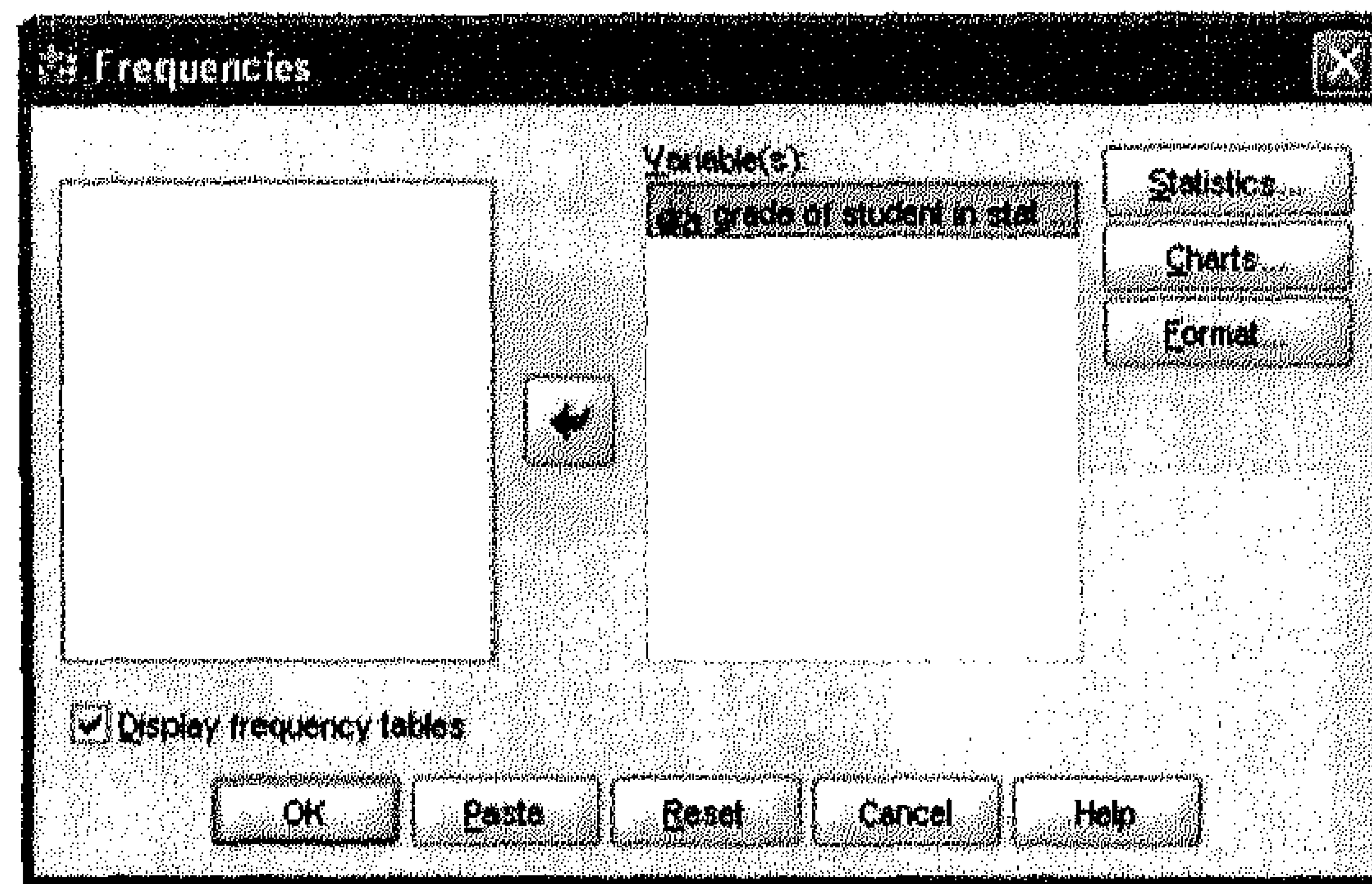
٥- نقوم بإدخال البيانات



٦- من قائمة Analyze نختار Descriptive Statistics فتظهر قائمة منسدلة نختار منها Frequencies

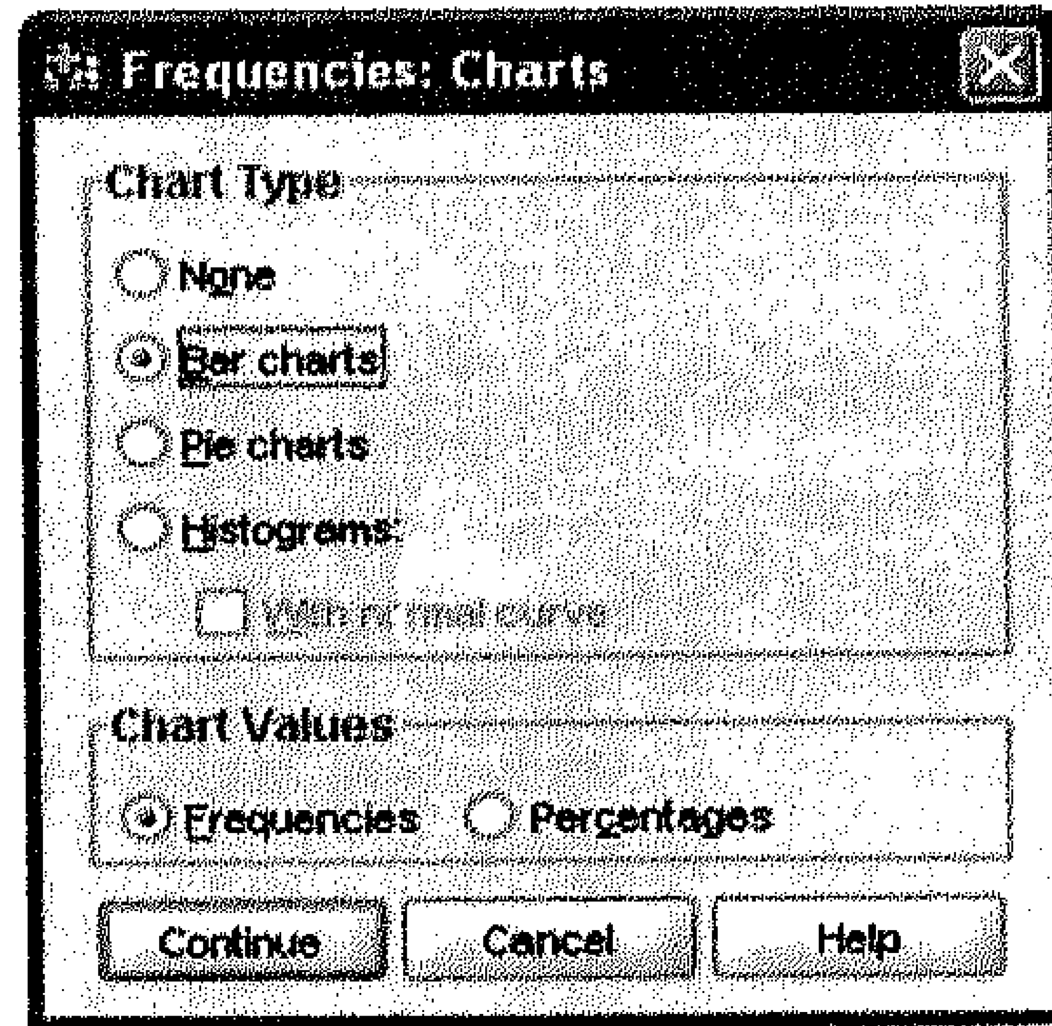


٧- تظهر شاشة جديدة بعنوان Frequencies ننقل المتغير grade إلى قائمة Variable(s) وذلك باختيار المتغير grade من القائمة الموجودة في يسار الشاشة ومن ثم نضغط على الزر ▶



نلاحظ الاختيار Display frequency tables ويستخدم لإنشاء الجدول التكراري عند اختياره

٨- فنضغط على الأمر Charts لتمثيل المتغير grade بيانياً ستظهر شاشة جديدة



نختار منها نوع التمثيل البياني الذي نرغب فيه ومن المعلوم أنه يمكن تمثيل البيانات الوصفية(النوعية) باستخدام كل من الأعمدة البيانية أو طريقة الدائرة البيانية:

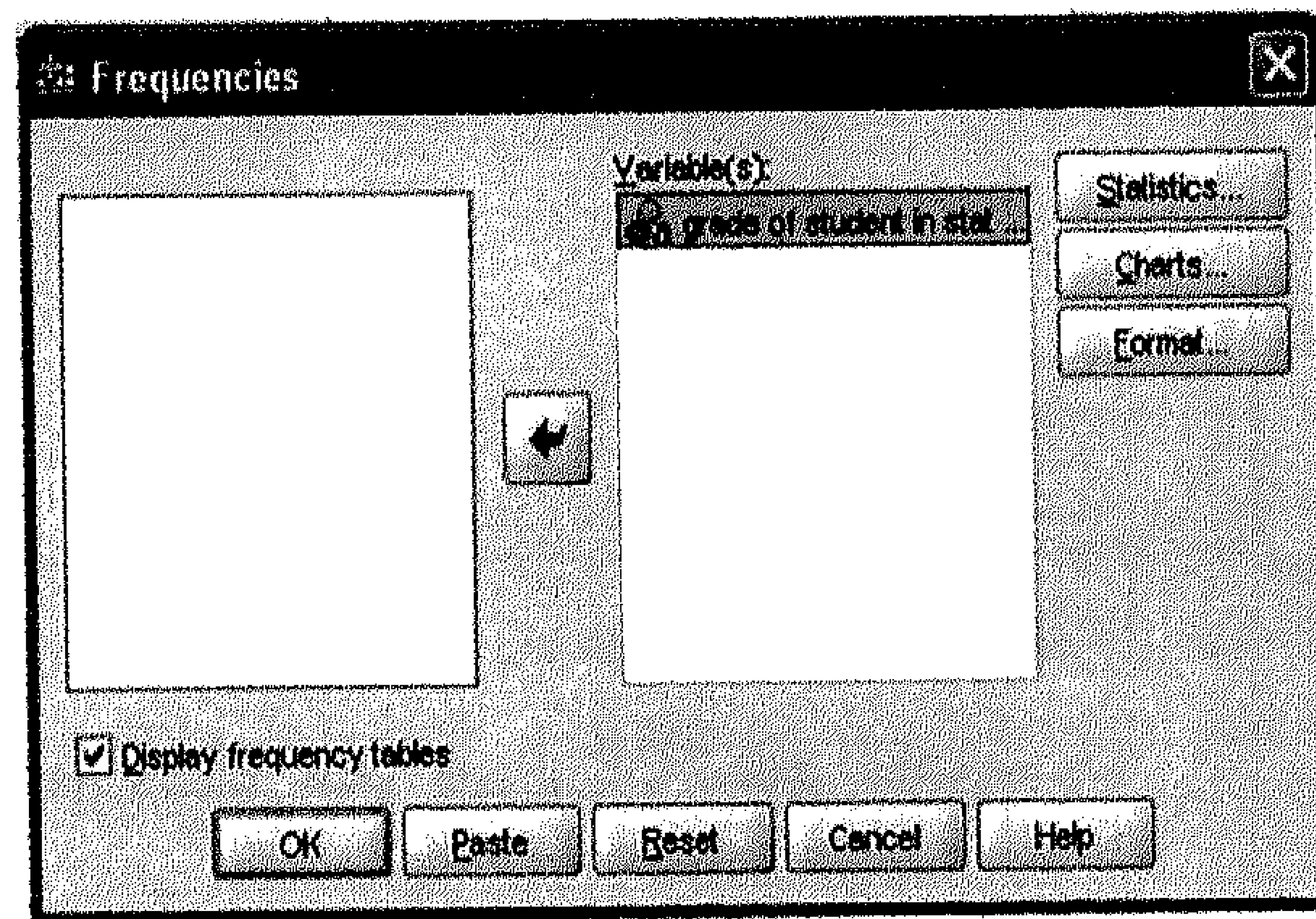
Bar charts: لاستخدام الأعمدة البيانية في التمثيل البياني

Pie Charts: لاستخدام الدائرة البيانية في التمثيل البياني

ونلاحظ أن البرنامج يتيح اختيار طريقة واحدة فقط من تلك المعروضة على الشاشة وسوف نختار Bar charts ثم نكرر فتح تلك الشاشة ونختار Pie charts

٩- ويوجد أسفل النافذة قائمة charts values والتي تتيح إجراء الرسم البياني عن طريق التكرارات Frequencies أو التكرار المئوي Percentages

١٠- نضغط على Continue فنعود للشاشة السابقة



١١- نختار Ok فتظهر شاشة المخرجات يحتوي على النواتج التالية:

الجدول الأول بعنوان Statistics ويعطى تلخيص عن اسم المتغير وعدد القيم للمتغير والقيم المفقودة

Statistics

grades to students in stat 1080

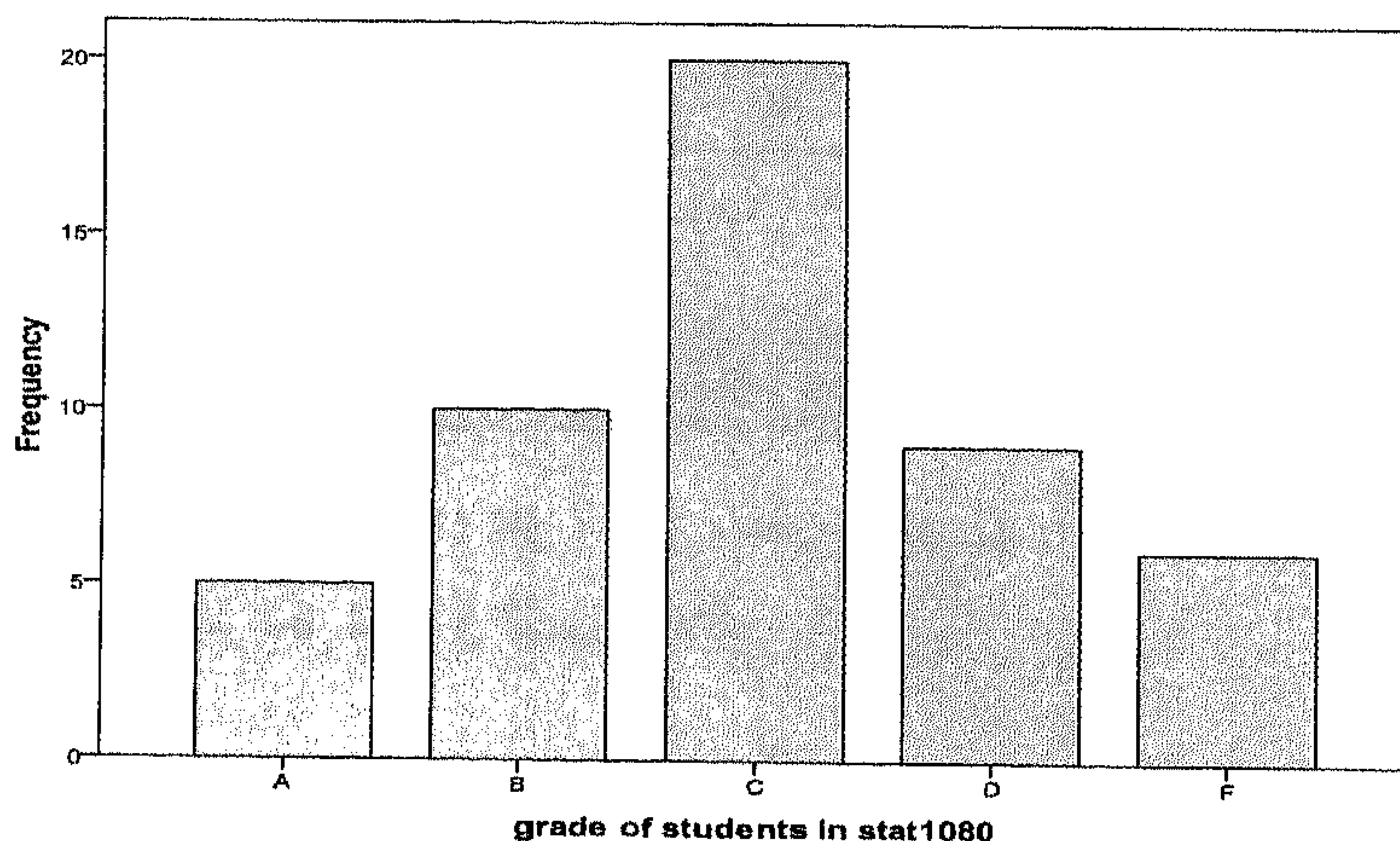
N	Valid	50
	Missing	0

الجدول الثاني يعطى الجدول التكراري للمتغير Grade وتظهر البيانات باستخدام Label الذي تم إضافته عند إدخال البيانات ويحتوى الجدول على قيم المتغير في العمود الأول والتكرار وأيضا التكرار المئوي والتكراري المئوي التراكمي.

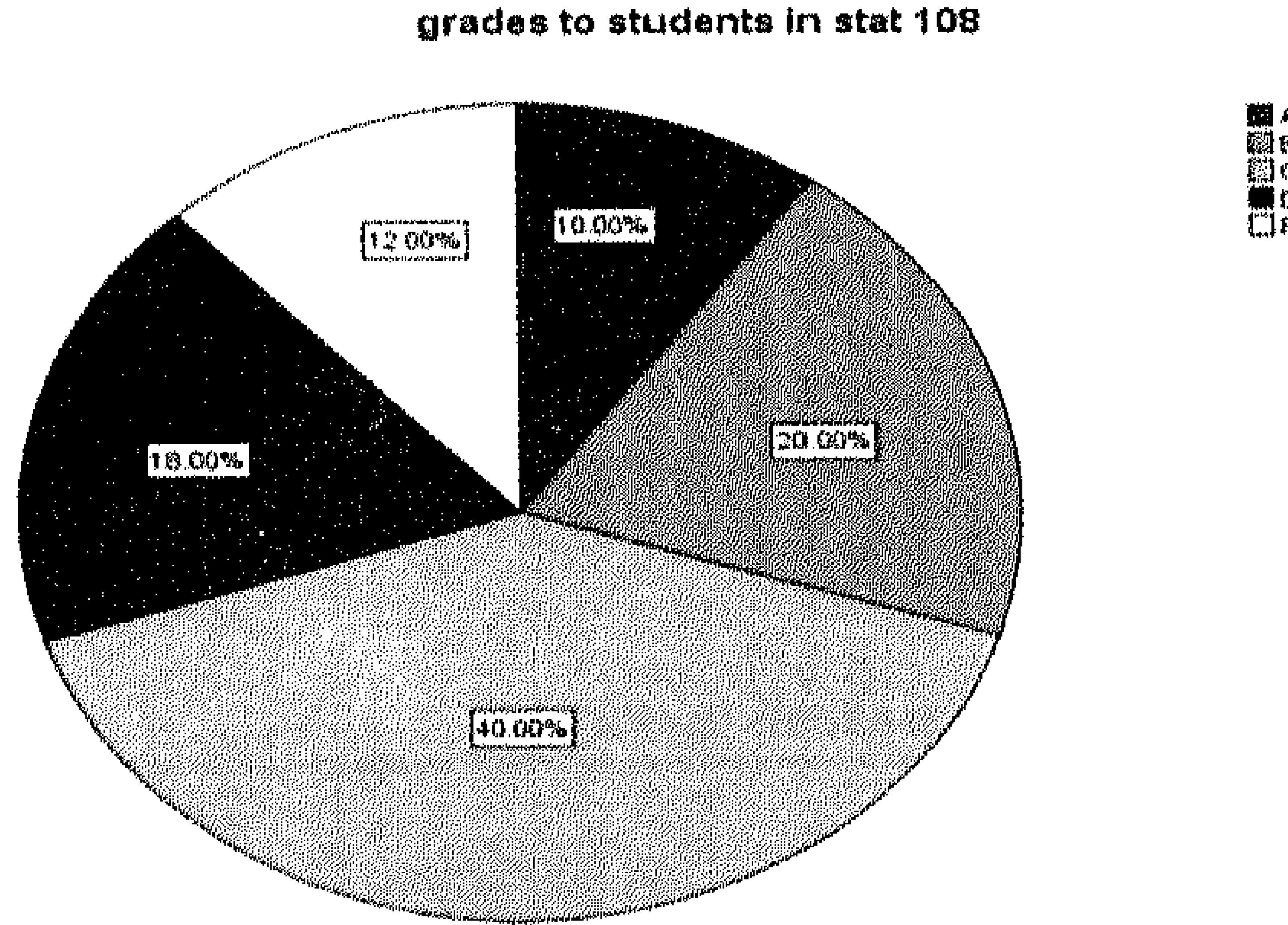
grades to students in stat 1080

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	A	5	10.0	10.0	10.0
	B	10	20.0	20.0	30.0
	C	20	40.0	40.0	70.0
	D	9	18.0	18.0	88.0
	F	6	12.0	12.0	100.0
	Total	50	100.0	100.0	

الشكل البياني وهو الأعمدة البيانية وقد قام باستخدام التكرار على المحور الرأسي (محور y) والقيم على المحور الأفقي



١٢- يمكن رسم الدائرة البيانية عن طريق تكرار ما سبق لكن في الخطوة ٨ سوف نختار Pie charts بدلاً من Bar charts وسوف نحصل على الشكل التالي:



تطبيق(٢-٢):

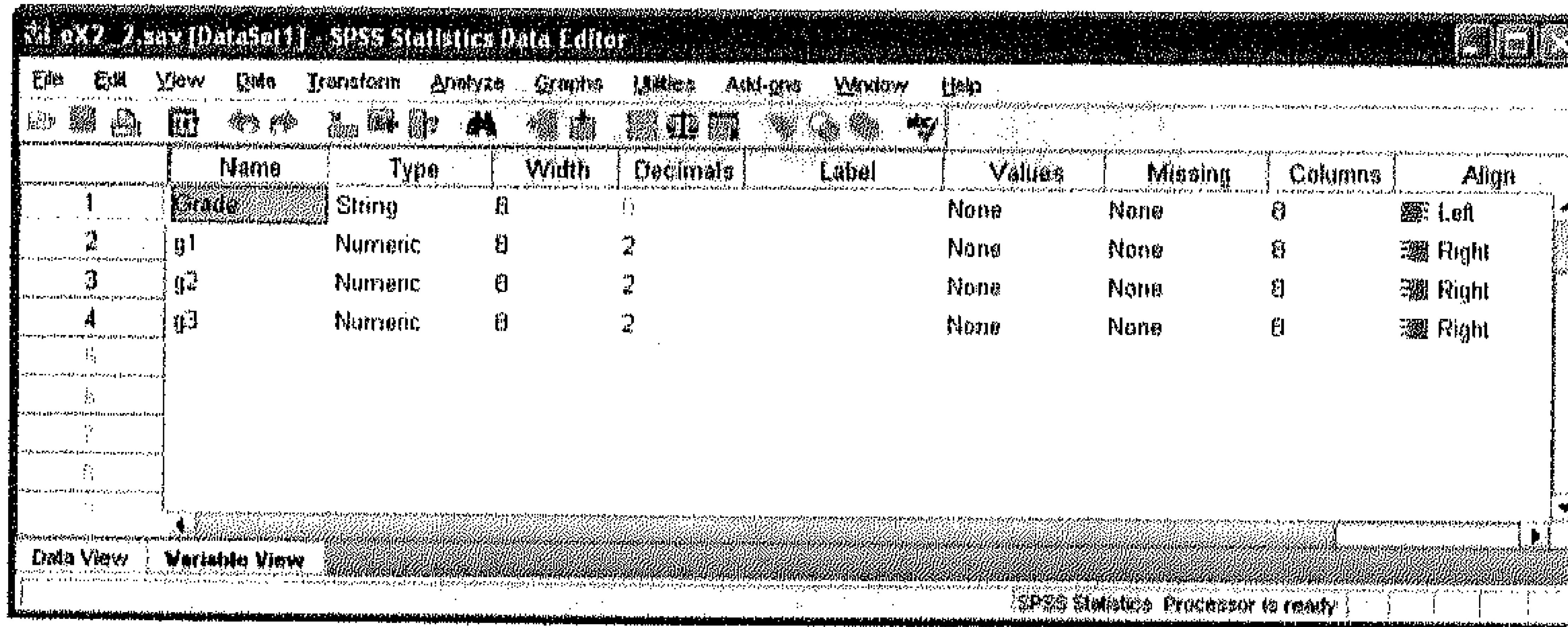
في مثال(٢-٢) إذا كان لدينا تقديرات الطلاب في ثلاثة شعب لمقرر 1080 إحص علماً بأن كل شعبة بها 40 طالباً وتم تلخيصها في الجدول التكراري التالي:

التقدير	تكرار(الشعبة الأولى)	تكرار(الشعبة الثانية)	تكرار (الشعبة الثالثة)
A	3	4	5
B	10	8	11
C	15	12	10
D	8	9	5
F	4	7	9
المجموع	40	40	40

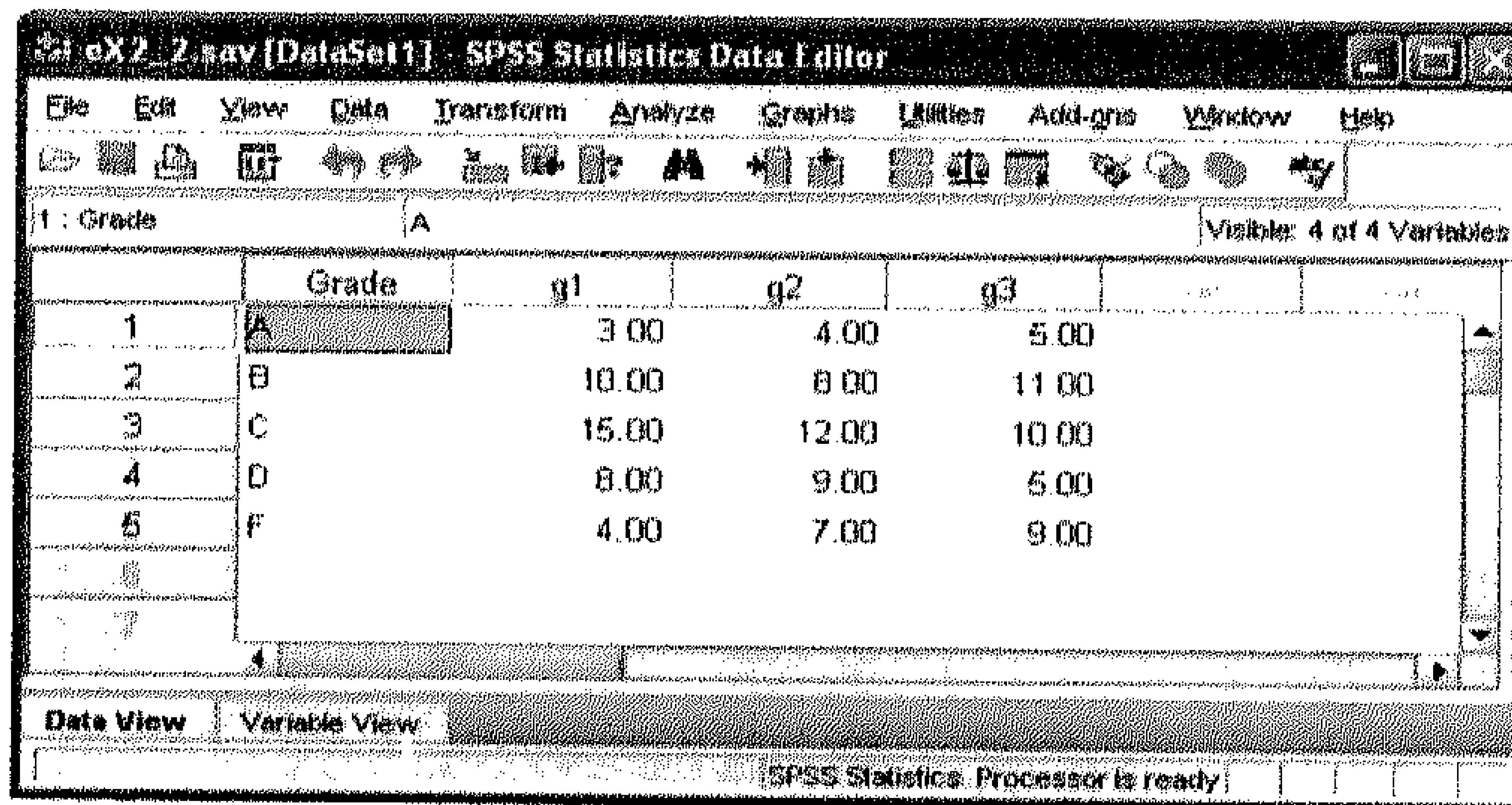
قارن بين الشعب الثلاثة باستخدام الأعمدة البيانية.

الحل:

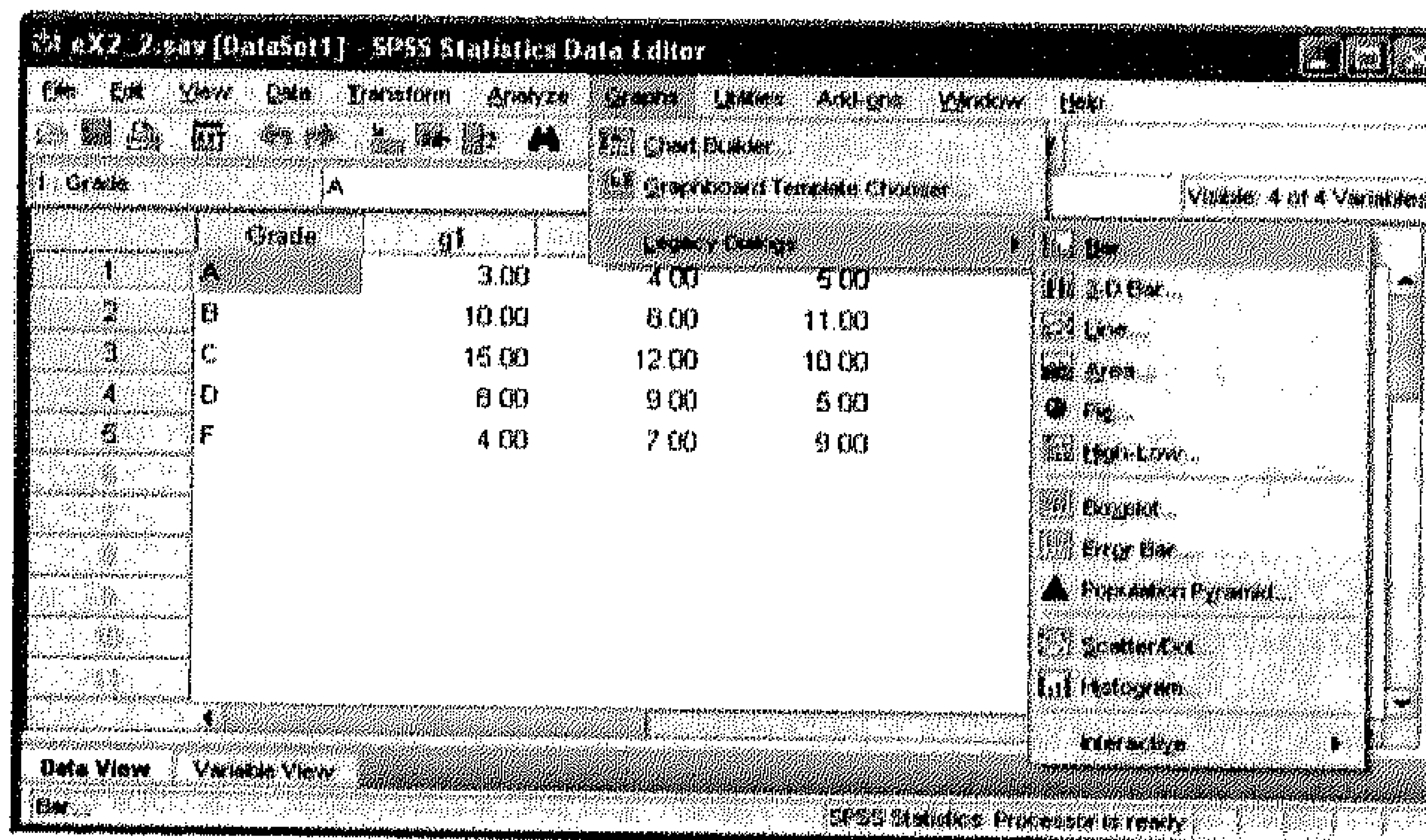
١- أولاً سنقوم بفتح ملف جديد وننتقل لنافذة Variable View ثم نقوم بتعريف أربعة متغيرات هي grade وهو متغير نوعى (string) يمثل التقدير والمتغيرات g1, g2, g3 وهى متغيرات عددية (Numeric) تمثل التكرار المناظر لكل تقدير في الشعب الثلاثة.



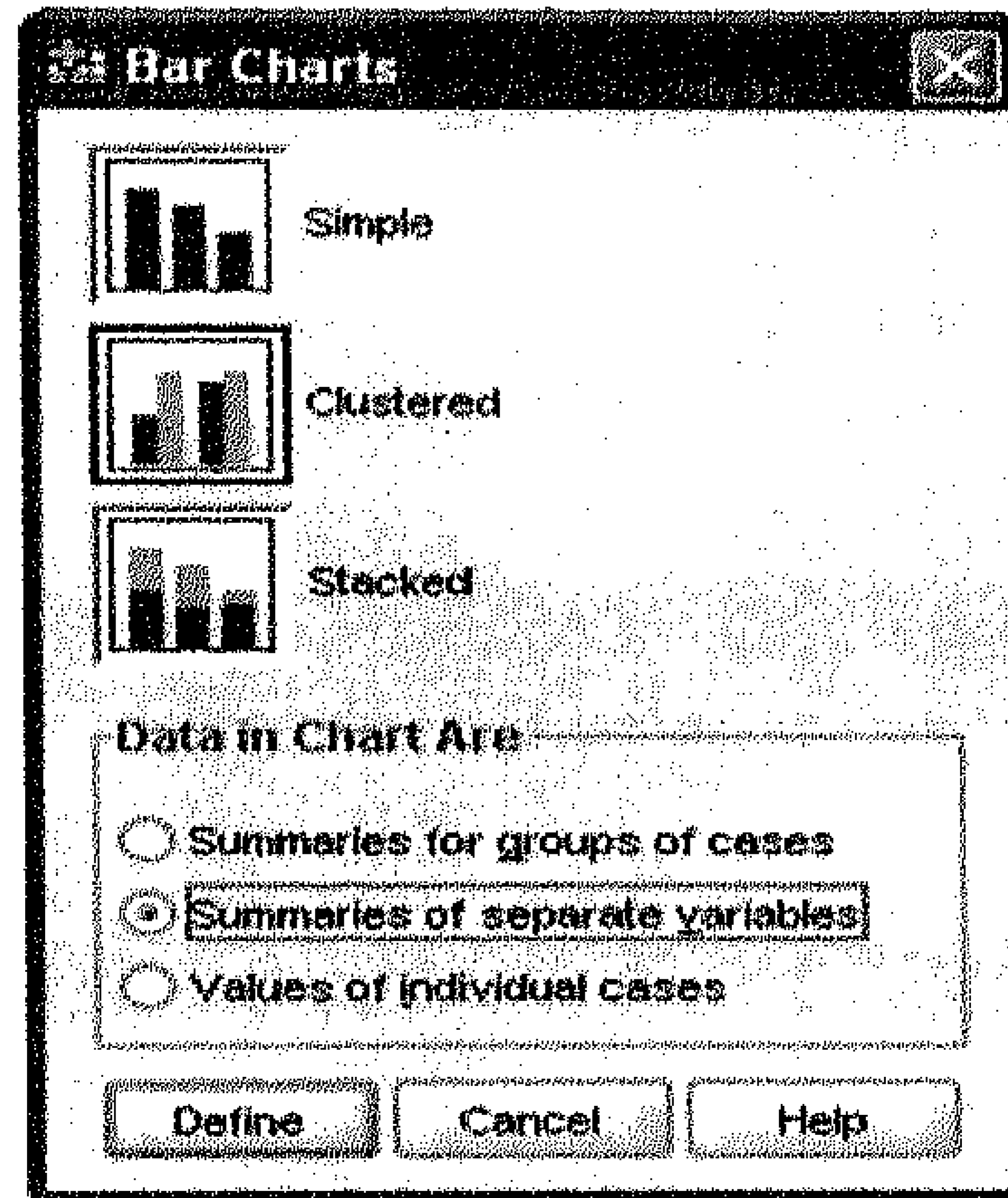
٢- ننتقل لنافذة Data View ونقوم بإدخال البيانات



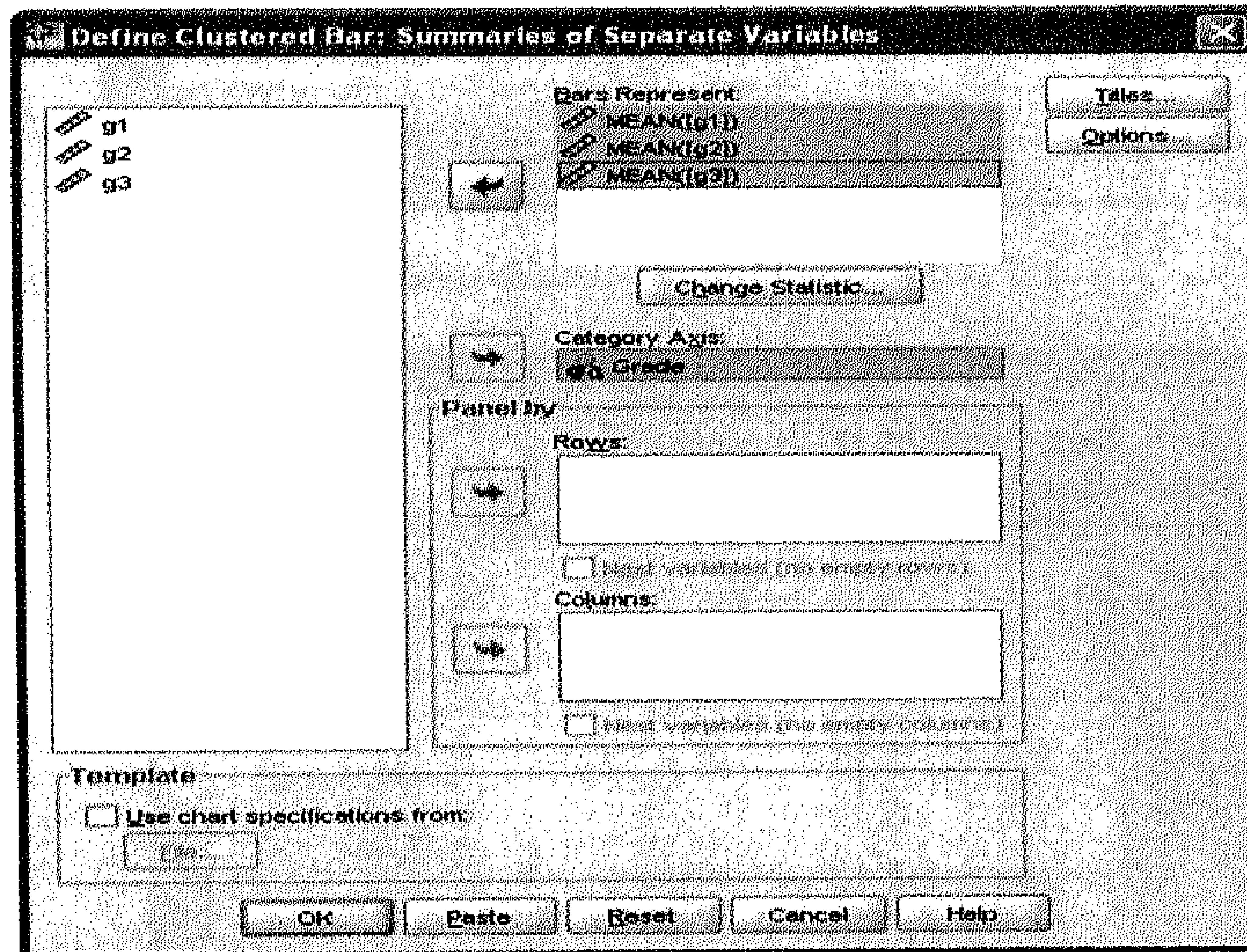
٣- من قائمة Graphs نختار Bar من القائمة المنسدلة Legacy Dialogs



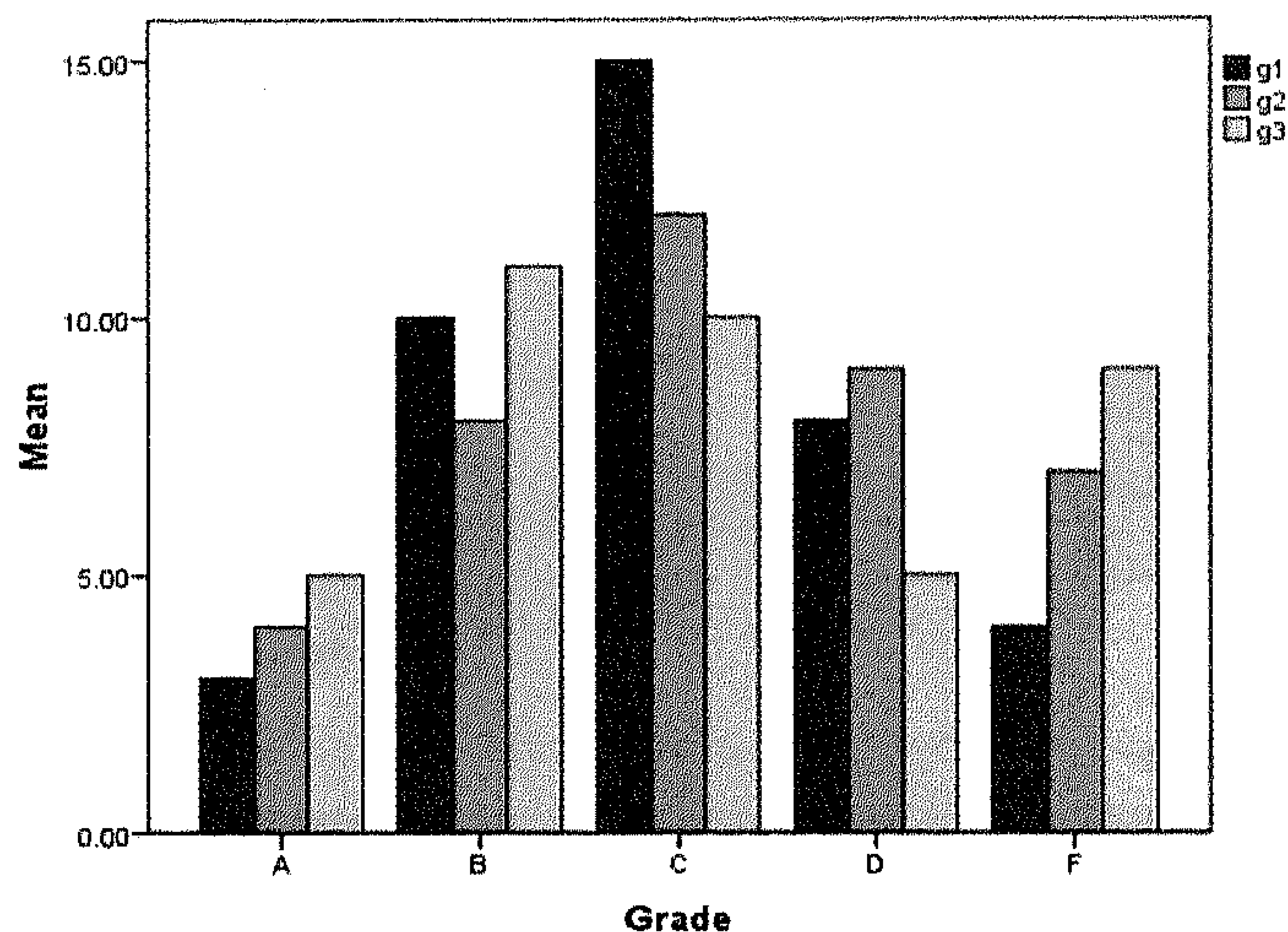
٤- فتظهر نافذة جديدة بعنوان Bar Charts



- نختار منها Clustered ومن قائمة Data in Charts Are نختار Summaries of separate variables ثم نضغط على Define فتظهر نافذة جديدة



- ننقل المتغير Grade لخانة Category Axis وننقل المتغيرات g1, g2, g3 لقائمة Bars Represent
- نختار Ok فتظهر النتائج التالية:



ثانياً المتغير العشوائي الكمي (Quantitative Variable)

١- المتغير الكمي المنفصل:

تطبيق(٢-٣):

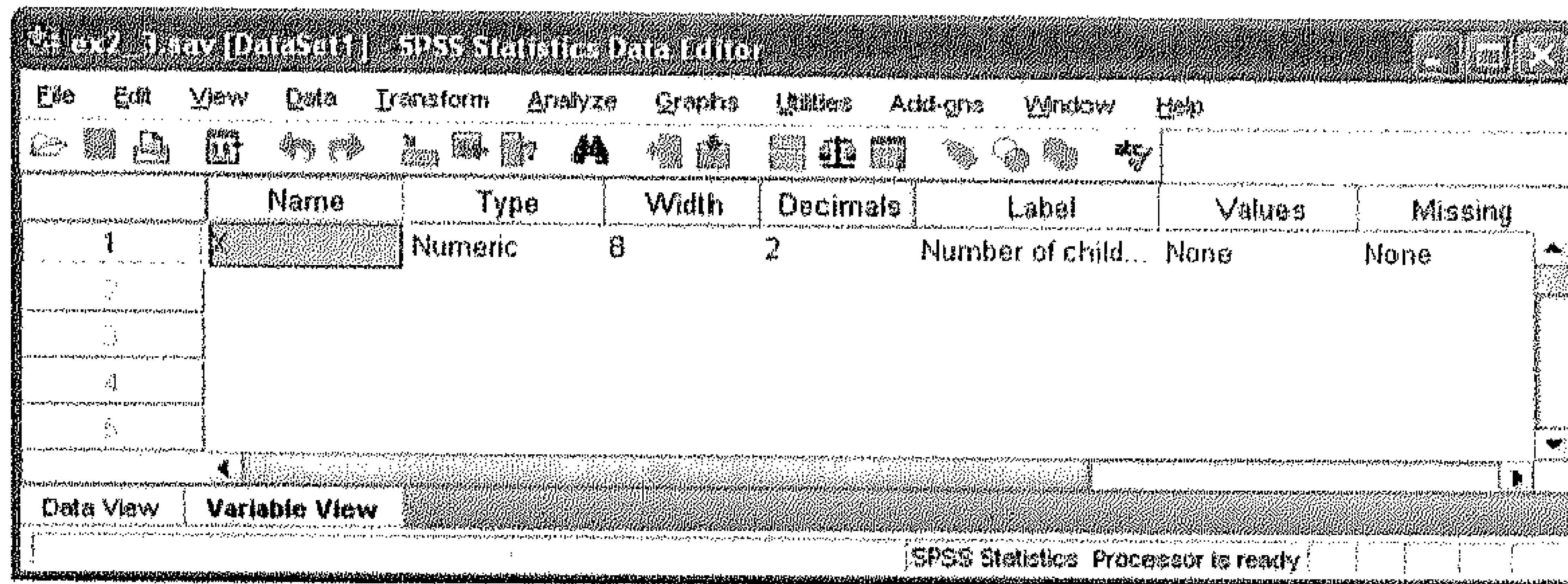
في مثال (٢-٤) في عينة من 30 أسرة تم اختيارها من إحدى المدن كان عدد الأطفال في كل أسرة كما يلي:

0	0	0	3	0	0	3	0
0	1	2	4	0	4	2	1
2	3	1	0	2	0	0	1

ضع هذه البيانات في صورة جدول تكراري ومثله بيانياً.

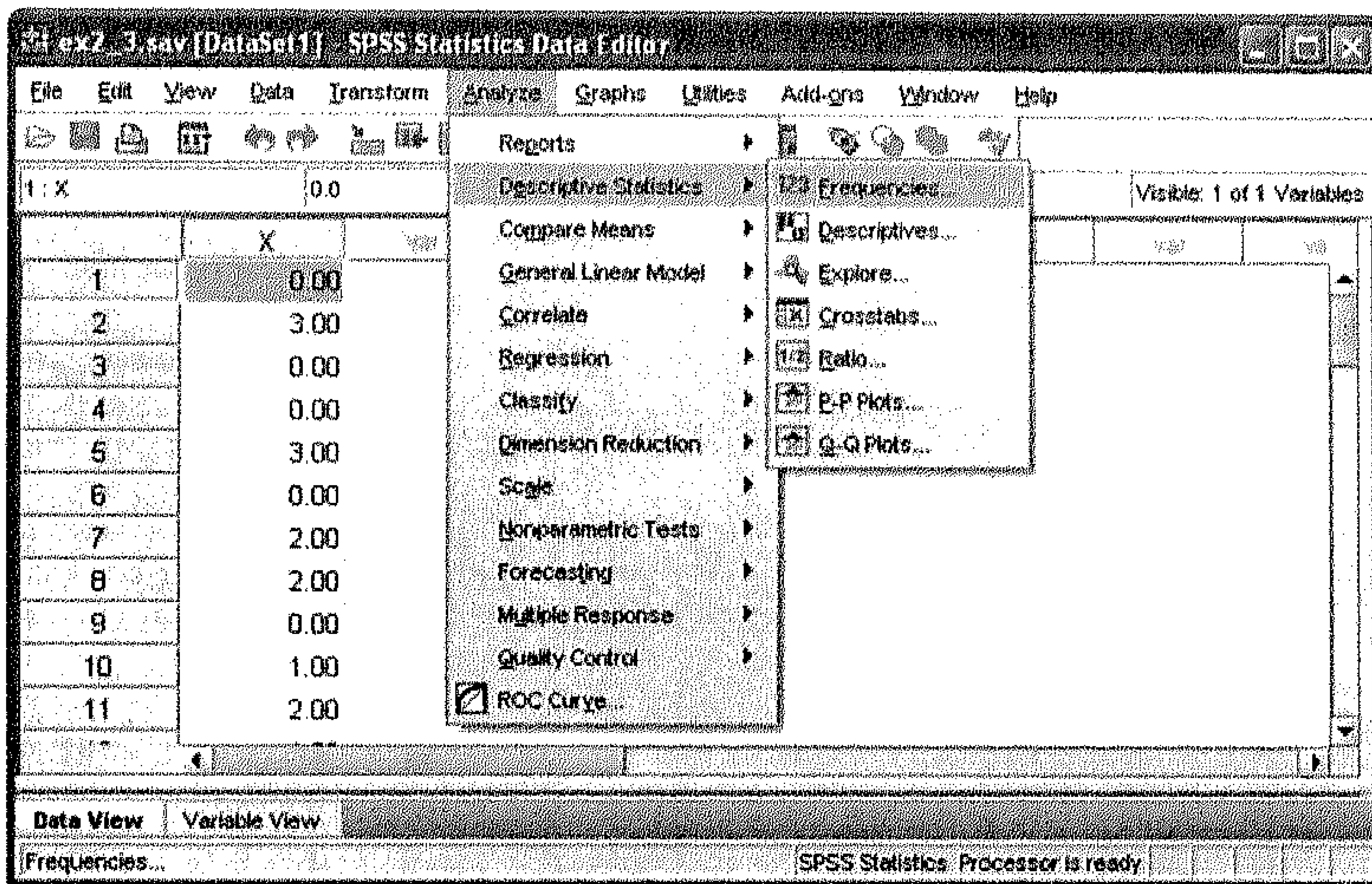
الحل:

١- سوف نقوم بتعريف المتغير X في نافذة Variable View على أنه Number of children in family وتحديد نوعه numeric

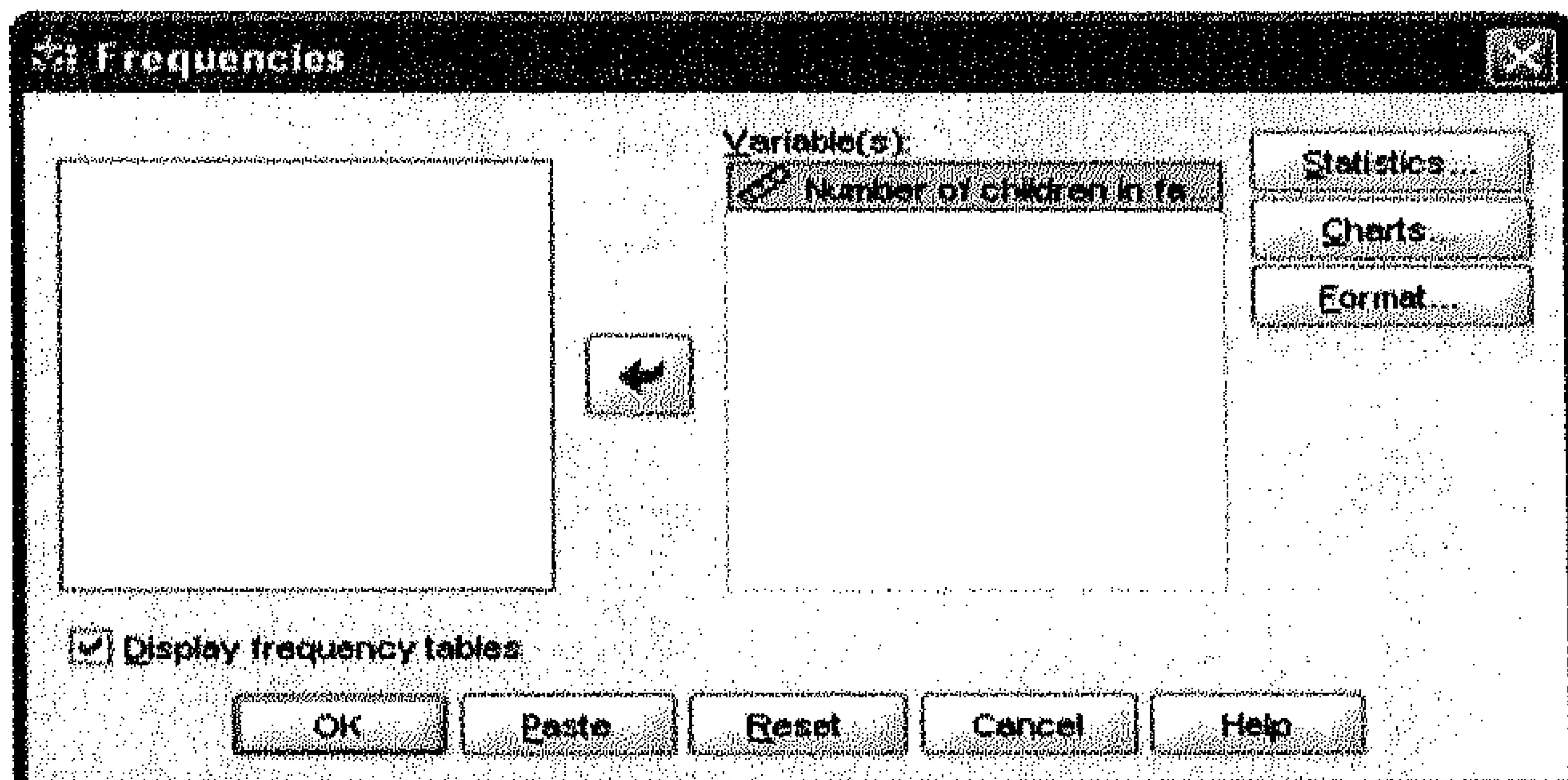


٢- بالانتقال لنافذة Data view نقوم بإدخال البيانات

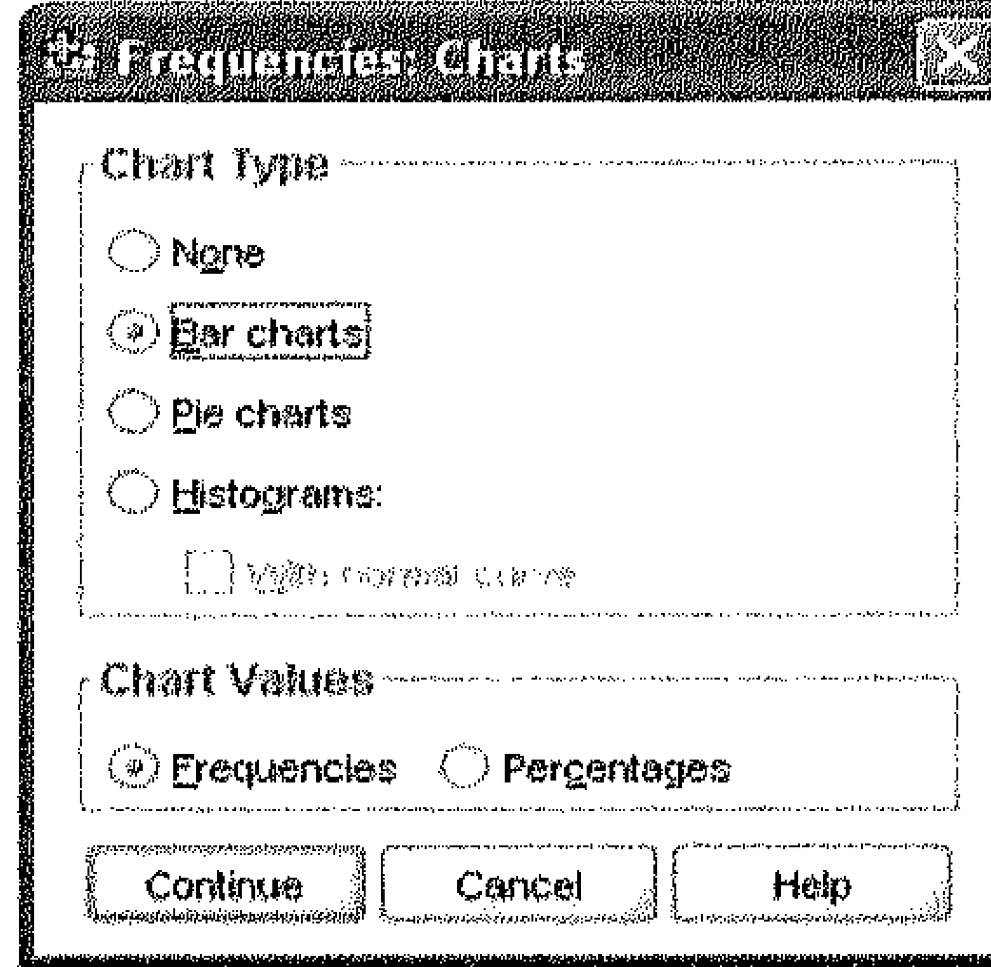
٣- من قائمة Analyze نختار Descriptive Statistics فتظهر قائمة منسدلة نختار منها Frequencies



٤- تظهر نافذة بعنوان Frequencies فننقل المتغير X لقائمة Variables



٥- نتأكد من أن الاختيار Display frequency tables محدد ثم نضغط على الأمر Charts فتظهر الشاشة التالية:



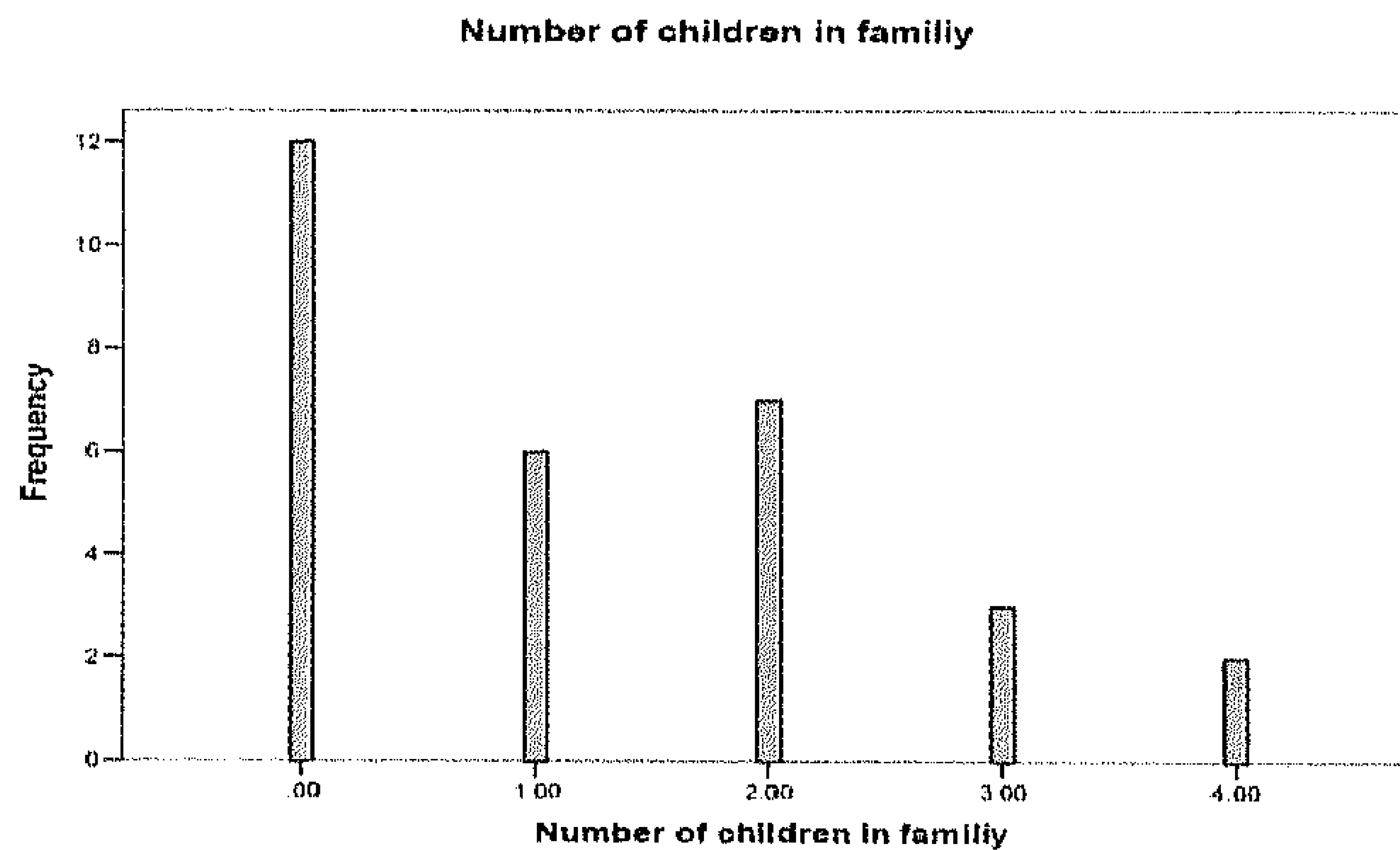
٦- نختار الأمر Bar charts ثم نختار Continue فنعود للشاشة السابقة ثم نضغط Ok فتظهر النواتج التالية:

- الجدول التكراري

Number of children in familiy

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00	12	40.0	40.0	40.0
	1.00	6	20.0	20.0	60.0
	2.00	7	23.3	23.3	83.3
	3.00	3	10.0	10.0	93.3
	4.00	2	6.7	6.7	100.0
	Total	30	100.0	100.0	

- الأعمدة البيانية



٢- المتغير الكمي المتصل:

تطبيق (٢-٤):

في مثال (٢-٦) البيانات التالية هي أوزان 57 طفل بالرطل في أحد مراكز الرعاية الطبية بالمملكة العربية السعودية:

68	63	42	27	30	23	36	28	32	79	46	27	16	24	69
22	23	24	25	44	19	65	43	25	74	30	51	47	23	49
36	42	28	31	28	43	25	45	12	57	12	51	22	43	27
12	32	49	38	42	49	27	31	50	38	28	21			

والمطلوب إنشاء الجدول التكراري لأوزان هؤلاء الأطفال، وحساب التكرار النسبي والمئوي لكل فترة وتمثيله بيانياً وذلك بفرض أن الحد الأدنى لأول فترة هو 10

الحل:

قبل التعامل مع برنامج SPSS يجب تحديد عدد الفترات في الجدول وطول كل منها حيث

عدد الفترات في الجدول التكراري هي:

$$k = 1 + 3.322 \times \log_{10}(n) = 1 + 3.322 \times \log_{10}(57) = 6.833 \cong 7$$

مدى البيانات هو:

$$R = X_{Max} - X_{Min} = 79 - 12 = 67$$

طول الفترة هو:

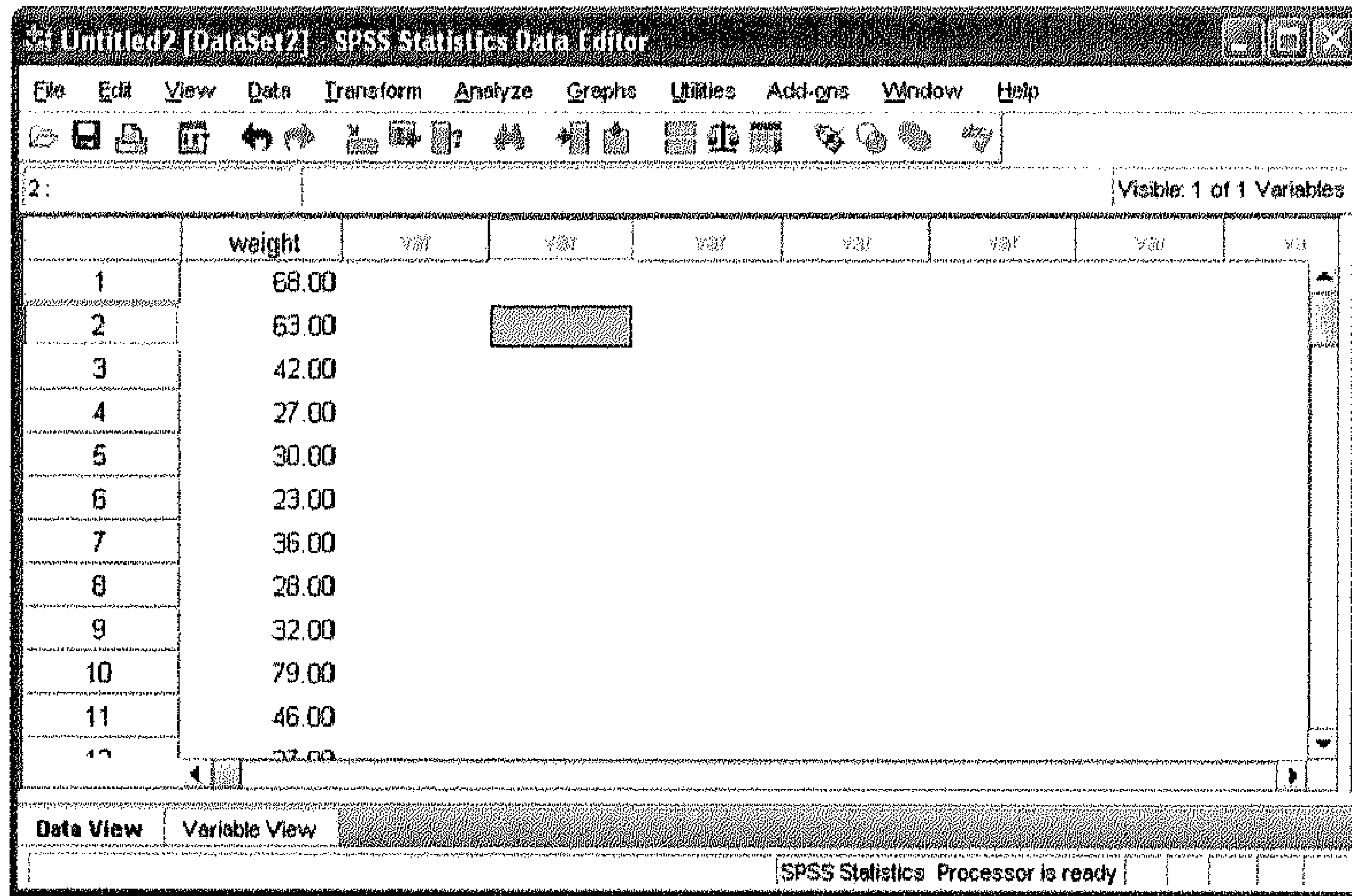
$$w = \frac{R}{k} = \frac{67}{7} = 9.57 \cong 10$$

الحدود التقريبية والفعلية للفترات هي:

الحدود التقريبية	الحدود الفعلية
10 – 19	9.5 – 19.5
20 – 29	19.5 – 29.5
30 – 39	29.5 – 39.5
40 – 49	39.5 – 49.5
50 – 59	49.5 – 59.5
60 – 69	59.5 – 69.5
70 – 79	69.5 – 79.5

١- من نافذة Variable view سنقوم بتعريف متغير عددي Numeric يمثل الأوزان Weight

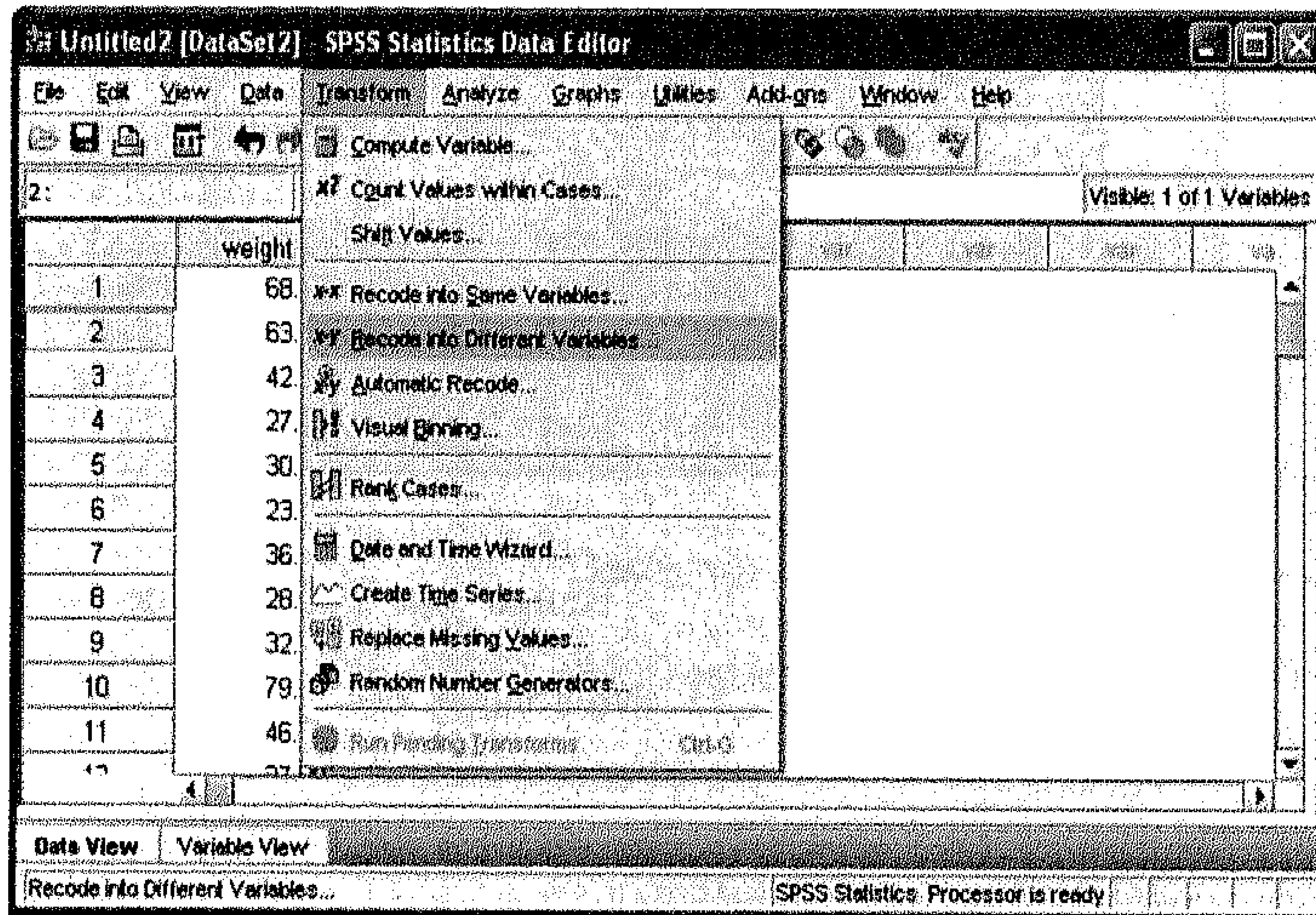
٢- ننتقل لنافذة Data View لإدخال البيانات كما في الشكل التالي:



٣- لإيجاد مراكز الفئات(الفترات) نقوم بالخطوات التالية:

يجب أولا إجراء عملية ترميز للبيانات

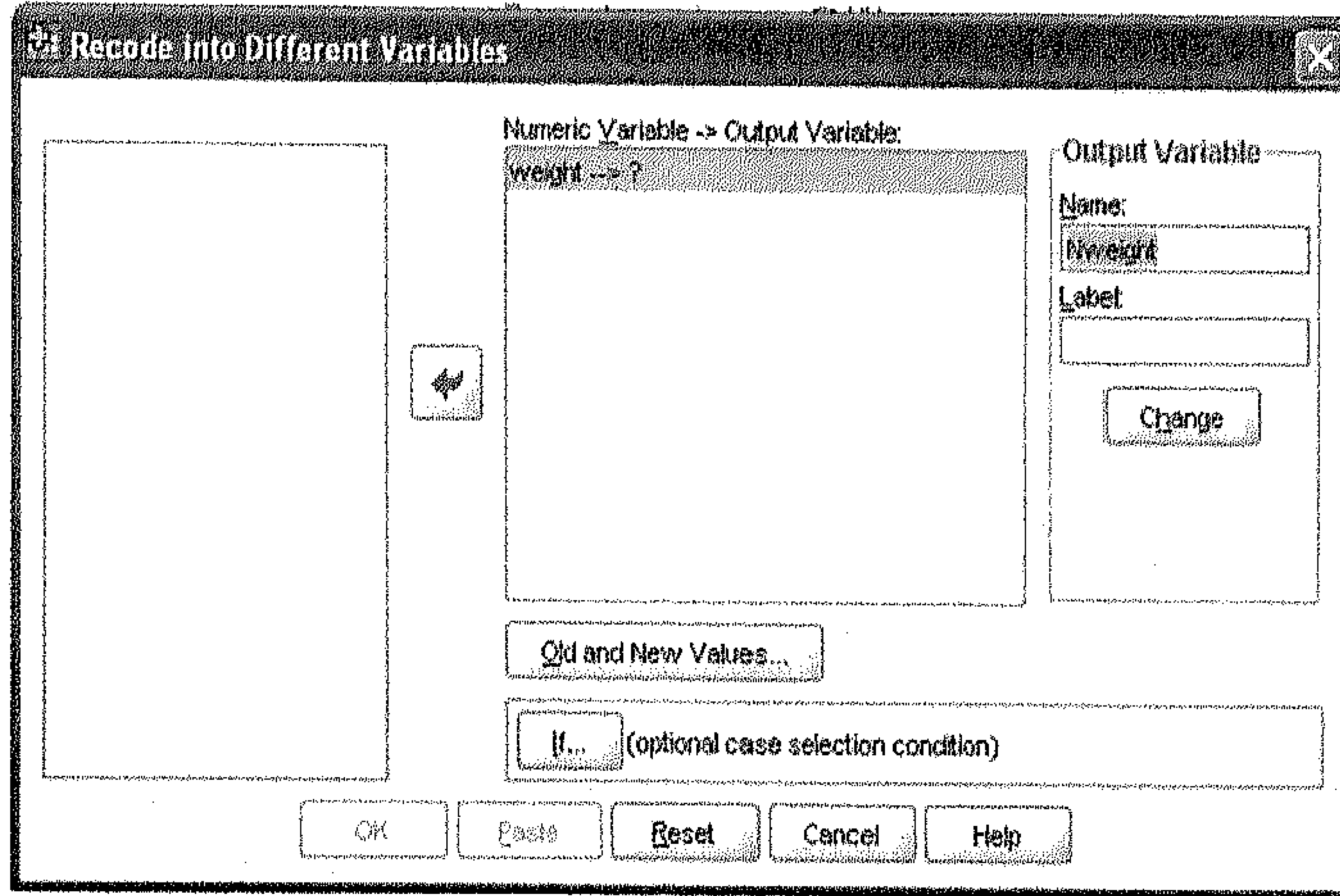
- من قائمة Transform نختار منها Recode into Different Variables...



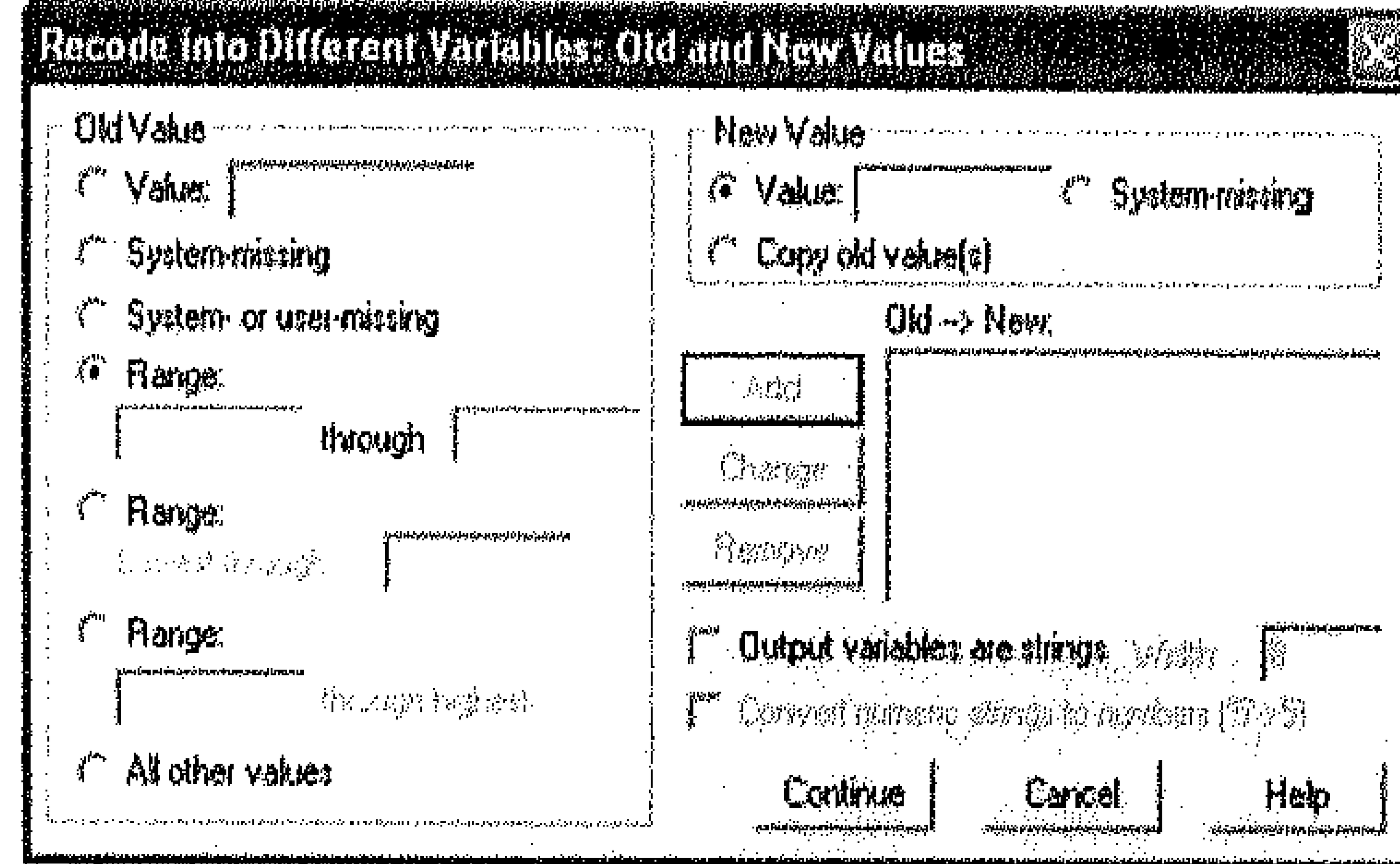
- سوف تظهر شاشة جديدة بعنوان Recode into Different Variables
- ننقل المتغير weight وفي خانة Output Variable نكتب الاسم الجديد الذي سيضاف فيه قيم المتغير

بعد عملية الترميز.

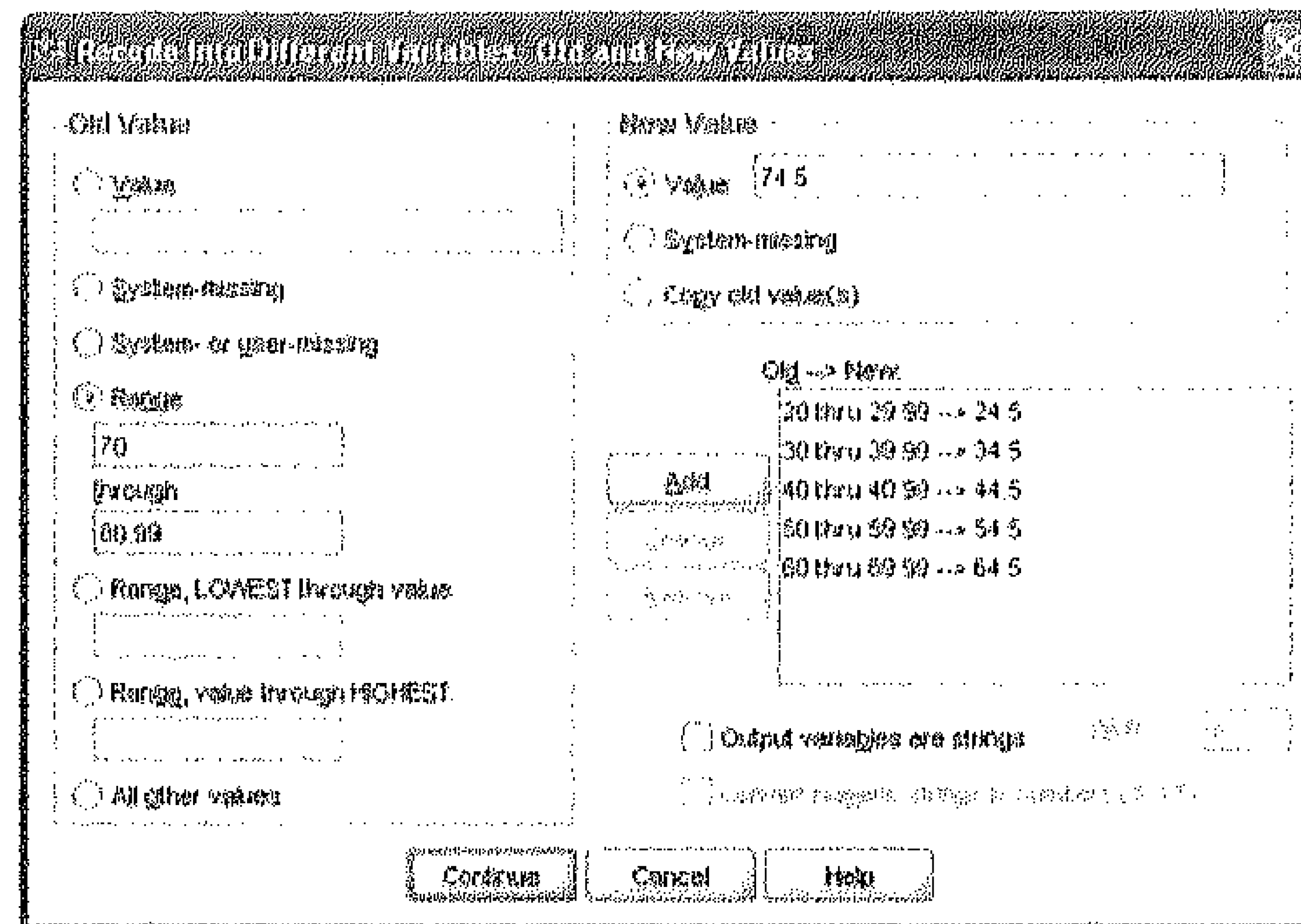
- ثم نضغط change



- نضغط على الاختيار Old and New Values فتظهر الشاشة التالية:



- نختار من قائمة Old Value الاختيار Range
- ثم نكتب حدود الفترات وفي خانة New value نكتب مركز الفترة ثم نضغط Add
- نكرر ذلك لكل الفترات



- ثم نضغط على Continue فنعود للشاشة السابقة
- نضغط على Ok نعود للملف الأصلي وقد قام بإضافة متغير جديد Nweight

Untitled2 [DataSet2] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

Visible: 2 of 2 Variables

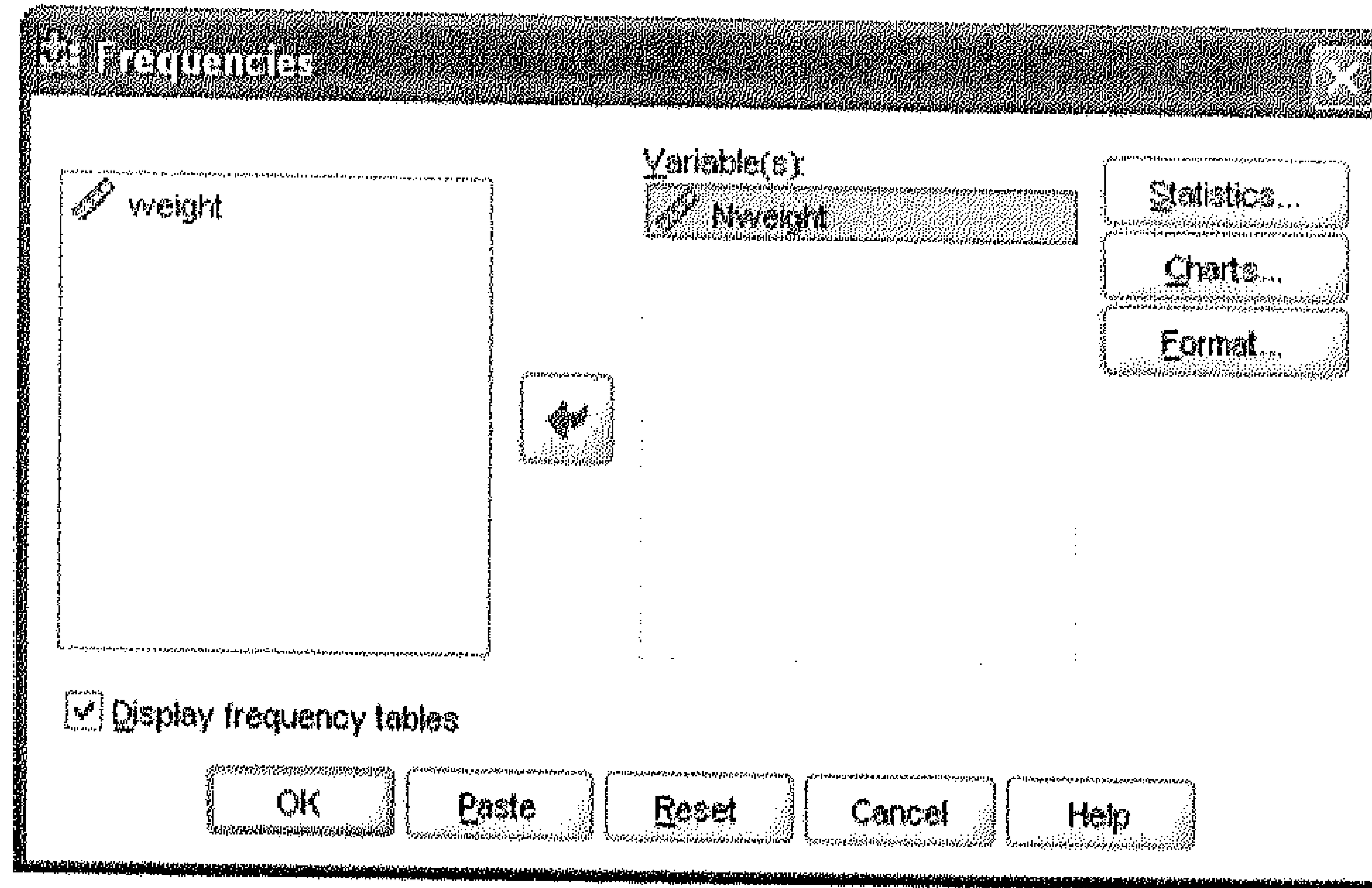
	weight	Nweight
1	68.00	64.50
2	63.00	64.50
3	42.00	44.50
4	27.00	24.50
5	30.00	34.50
6	23.00	24.50
7	36.00	34.50
8	28.00	24.50
9	32.00	34.50
10	79.00	74.50
11	46.00	44.50

Data View | Variable View

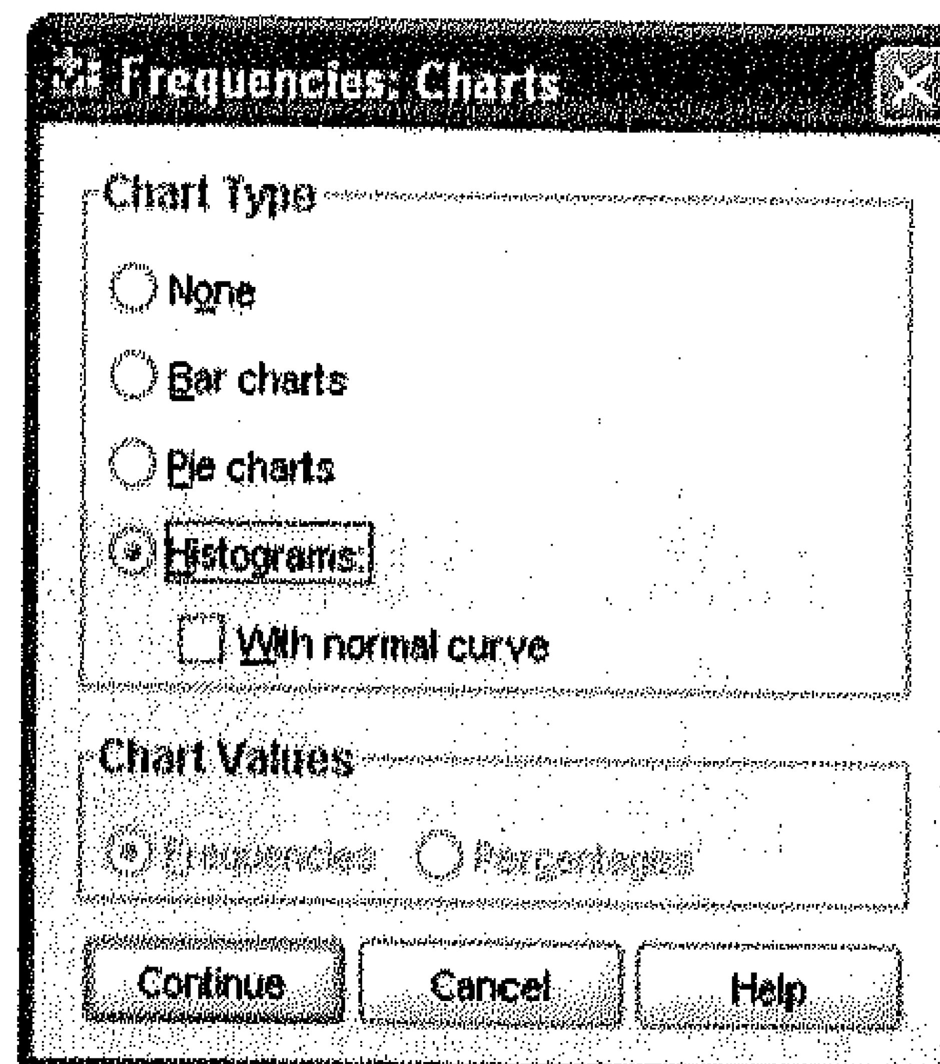
SPSS Statistics Processor is ready

٤- لإنشاء الجدول التكراري والتمثيل البياني باستخدام الأعمدة البيانية سوف نقوم باتباع التالي:

- من قائمة Analyze نختار Descriptive Statistics
- فتظهر قائمة منسدلة نختار منها Frequencies فتظهر شاشة جديدة بعنوان Frequencies



- ننقل المتغير Nweight لقائمة Variable(s) ثم نضغط على Charts
- نختار Histogram من النافذة الجديدة



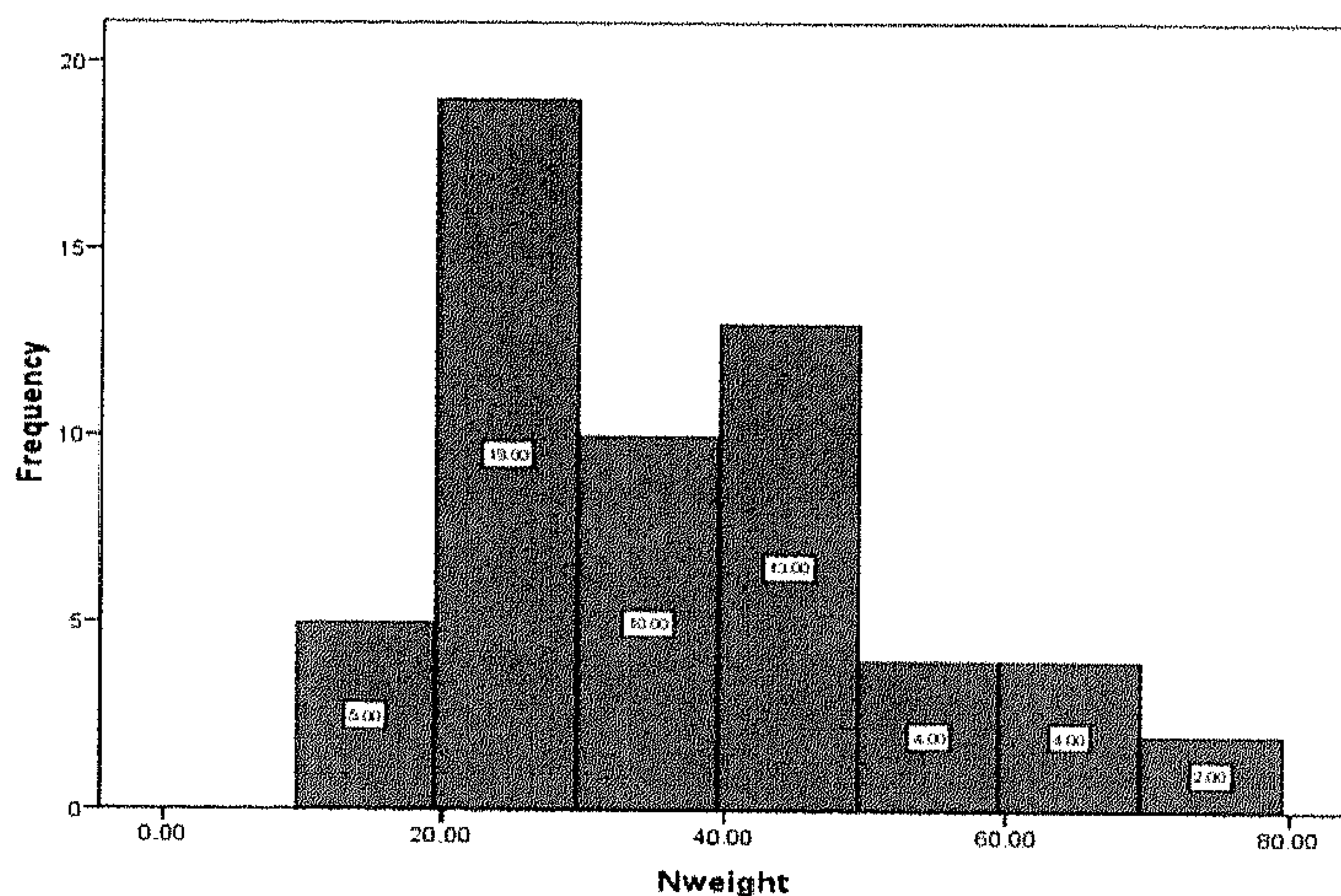
- نضغط على Continue فنعود للشاشة السابقة ونتأكد أن الاختيار Display frequency tables نشط ثم نختار Ok فتظهر النتائج التالية:

الجدول التكراري يحتوى العمود الأول فيه على مركز الفترات ثم العمود الثاني على التكرار، ثم التكرار المئوي والعمود الأخير التكرار المئوي التراكمي (المتجمع الصاعد)

Nweight

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	14.50	5	8.8	8.8	8.8
	24.50	19	33.3	33.3	42.1
	34.50	10	17.5	17.5	59.6
	44.50	13	22.8	22.8	82.5
	54.50	4	7.0	7.0	89.5
	64.50	4	7.0	7.0	96.5
	74.50	2	3.5	3.5	100.0
	Total	57	100.0	100.0	

الرسم البياني يمثل الأعمدة البيانية.



ولرسم المضلع التكراري أو المنحنى التكراري يجب إدخال البيانات على الصورة التالية:

التكرار (f)	مركز الفترات (X)
0	4.5
5	14.5
19	24.5
10	34.5
13	44.5
4	54.5
4	64.5
2	74.5
0	84.5
57	المجموع

وسوف نضيف فترة قبل الفترات الموجودة وبعدها ويكون تكرارها صفر حتى يكون المضلع التكراري مغلقاً.

- نقوم بإدخال البيانات في صورة متغيرين عدديين

*ex2_5p.sav [DataSet3] - SPSS Statistics Data Editor

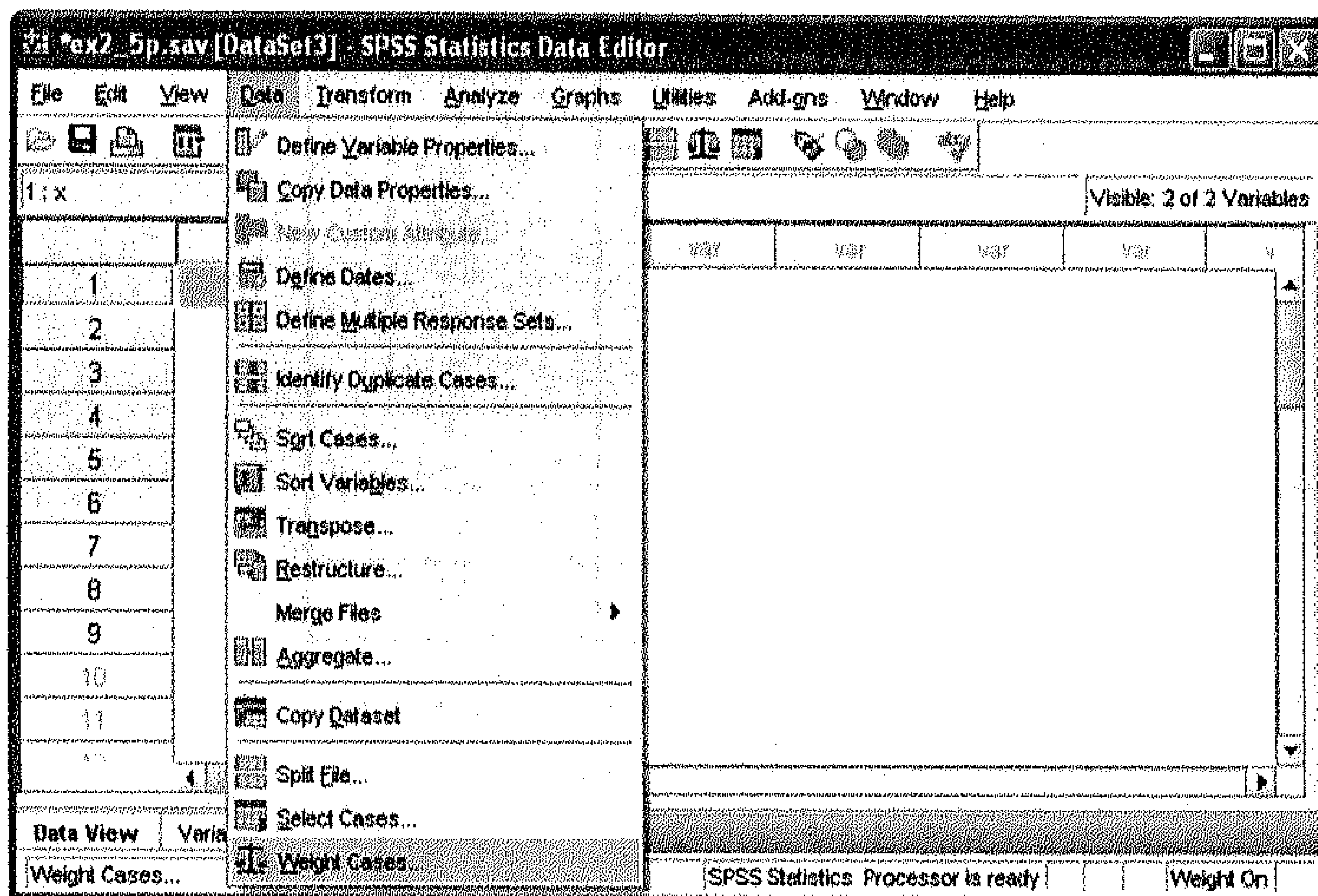
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : x 4.5 — Visible: 2 of 2 Variables

	x	f
1	4.50	0.00
2	14.50	5.00
3	24.50	19.00
4	34.50	10.00
5	44.50	13.00
6	54.50	4.00
7	64.50	4.00
8	74.50	2.00
9	84.50	0.00

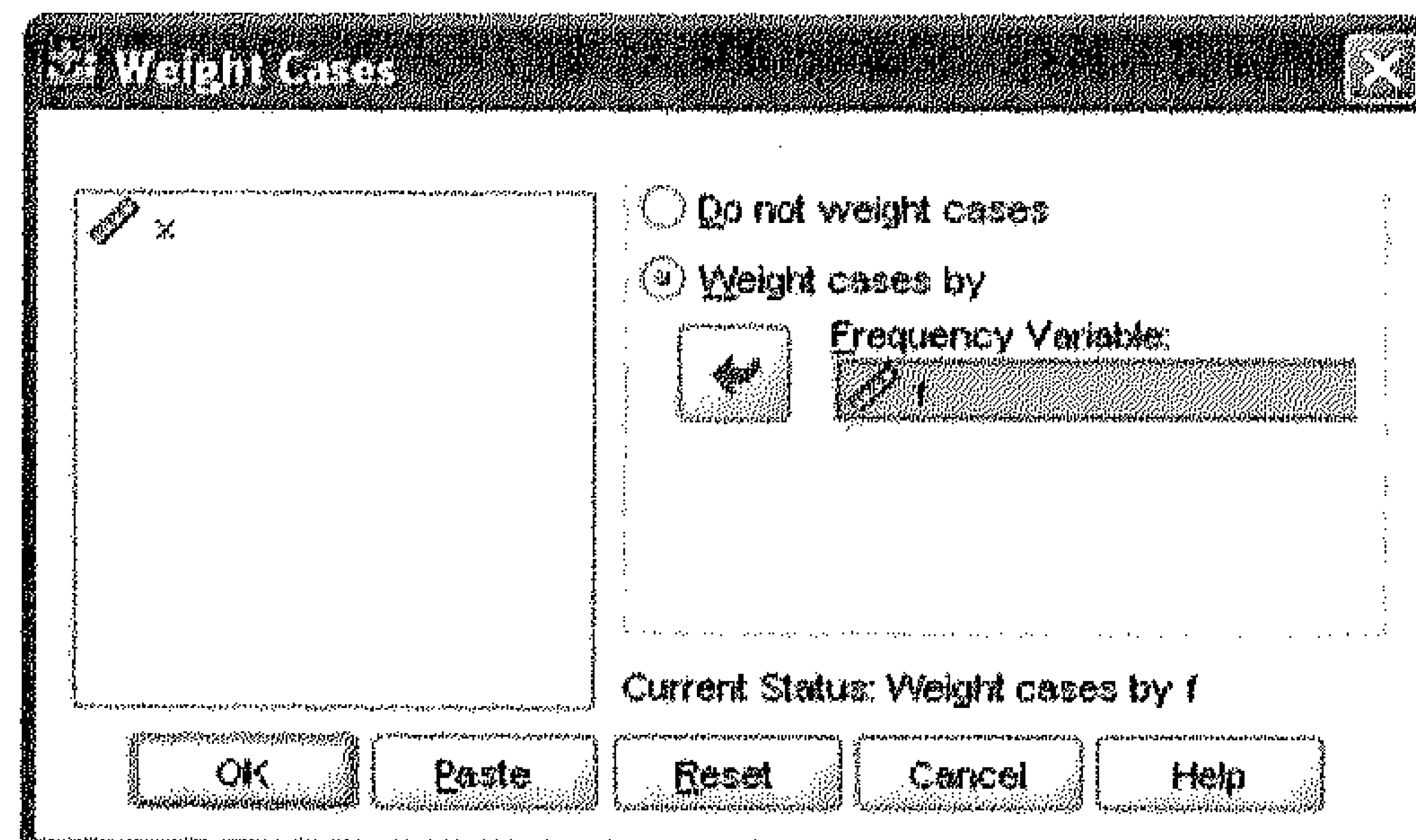
Data View | Variable View — SPSS Statistics Processor is ready — Weight On

● يجب أن نحدد أن f هي تكرار للفئة التي مركزها x وذلك باستخدام الأمر weight cases من قائمة Data

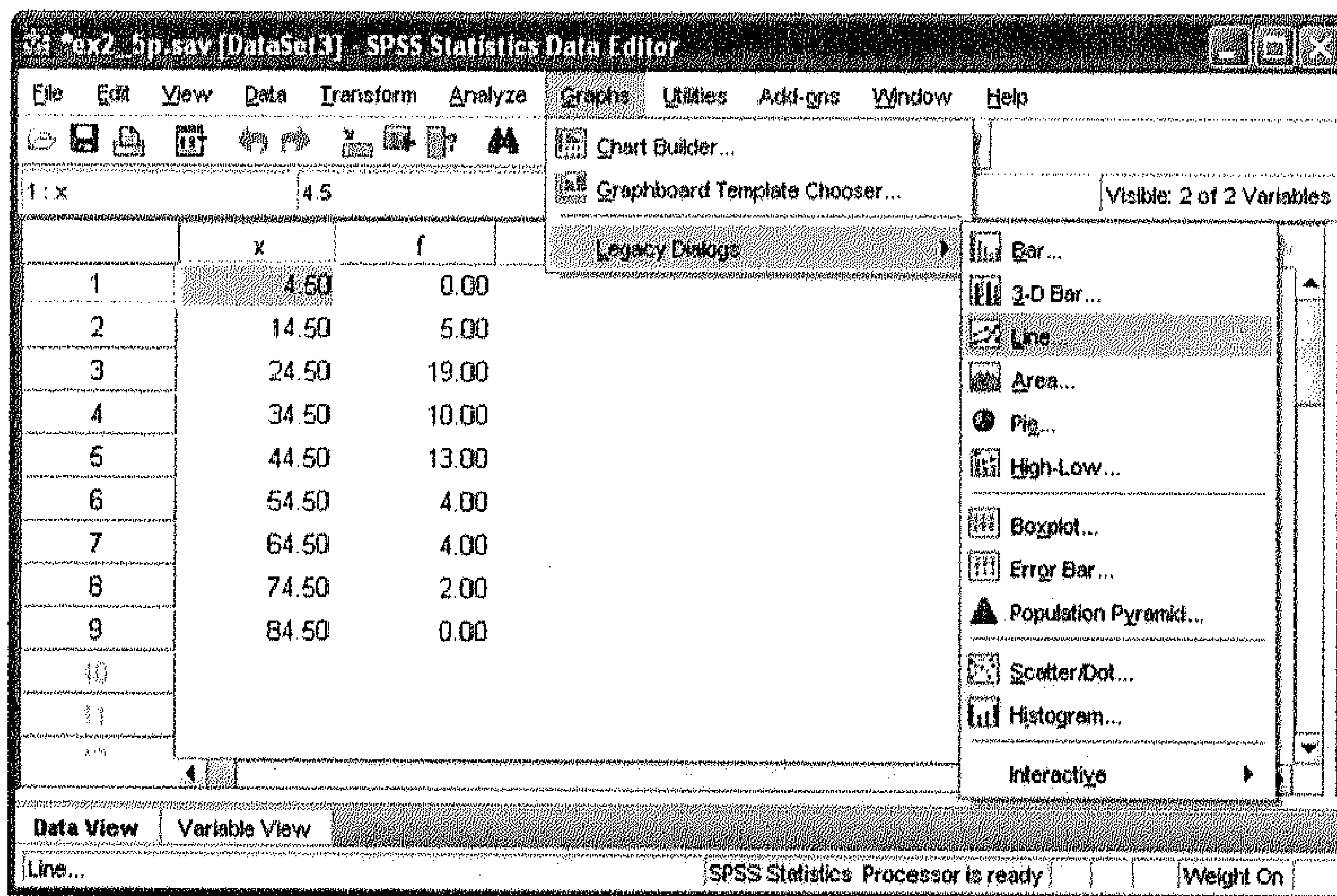


● سوف تظهر شاشة جديدة نختار منها Weight cases by ثم ننقل المتغير f لخانة Frequency Variable

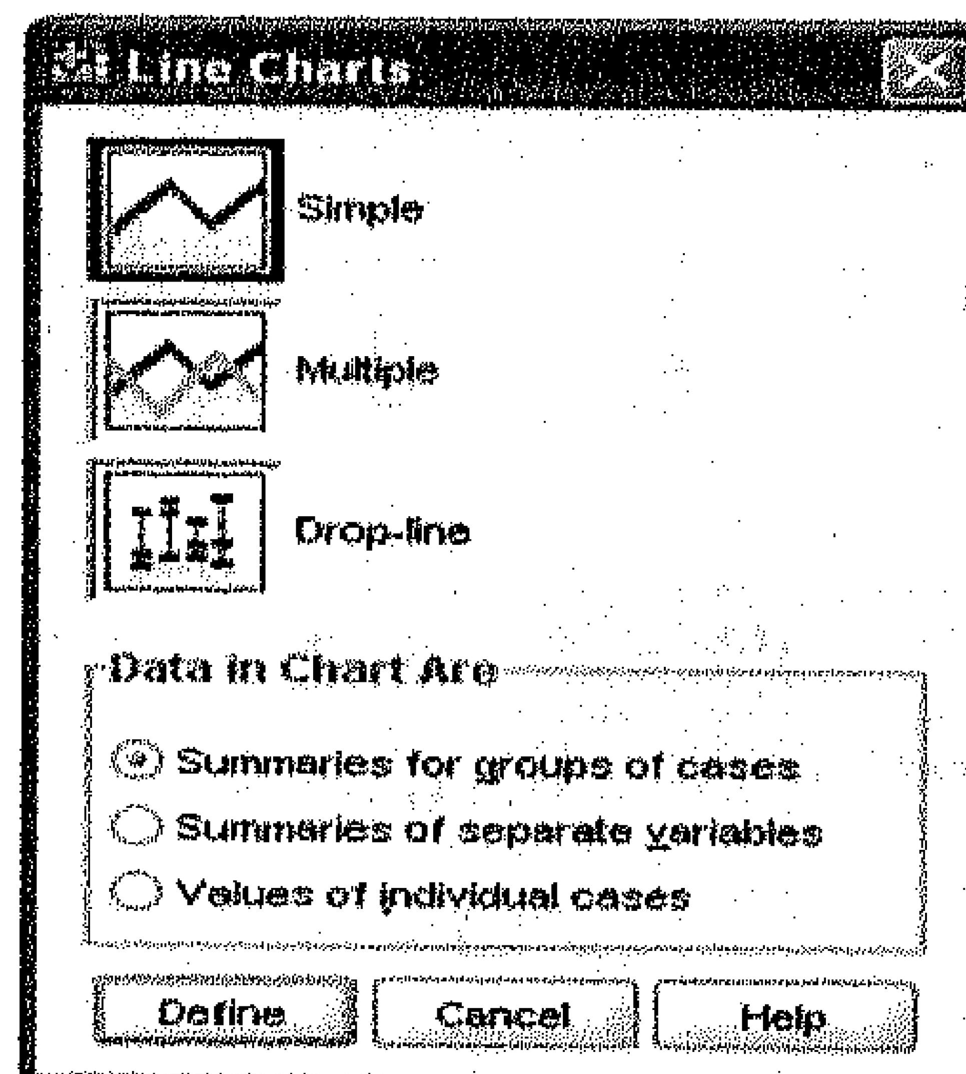
● ثم نختار Ok فنعود إلي ملف البيانات الأصلي ونلاحظ أن الملف لم يحدث له أي تغيير ظاهري.



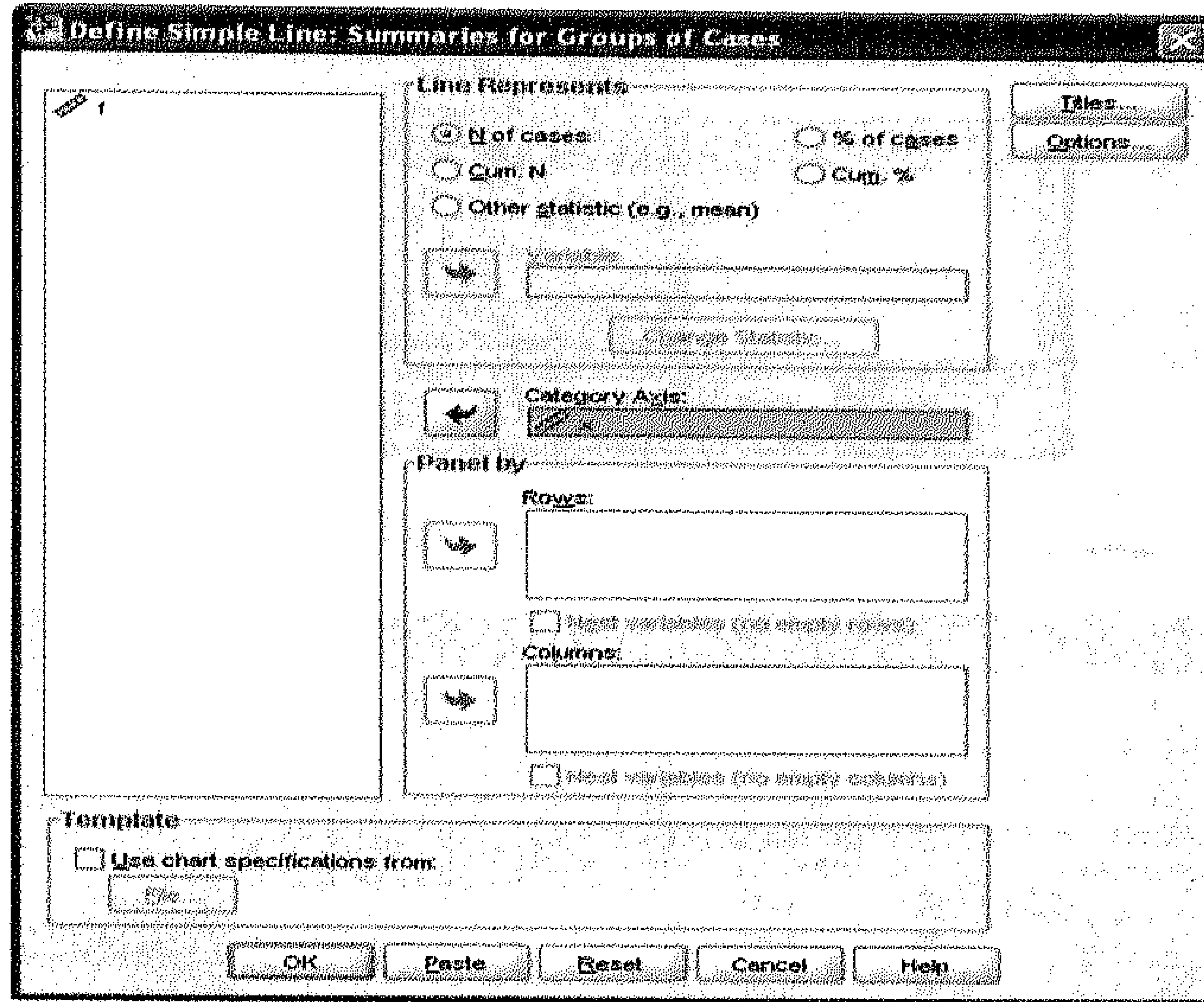
- نضغط على قائمة Graphs من القائمة المنسدلة Legacy Dialogs نختار Line فتظهر شاشة جديدة



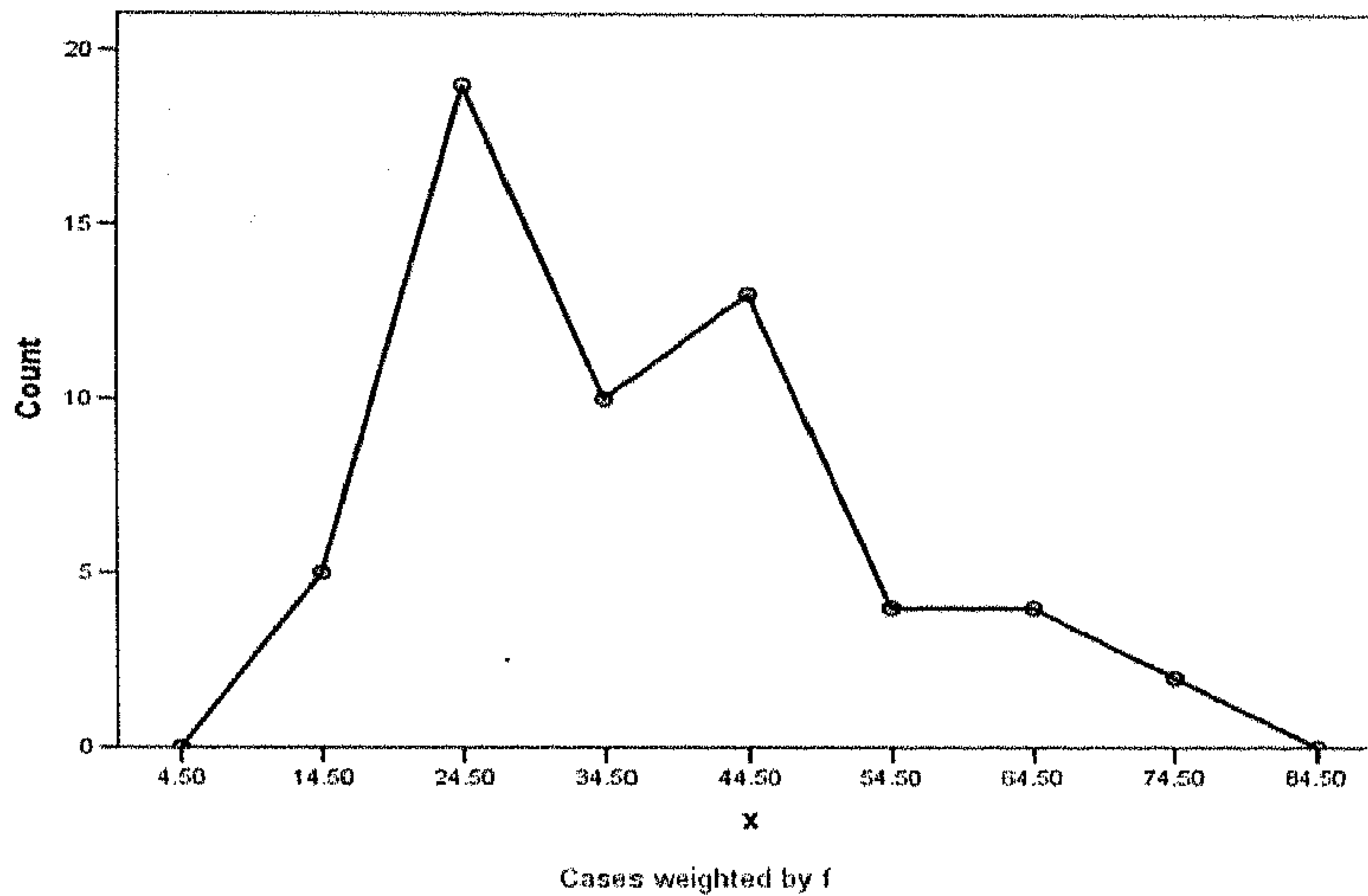
- من نافذة Line Charts نختار Simple ثم نضغط على Define



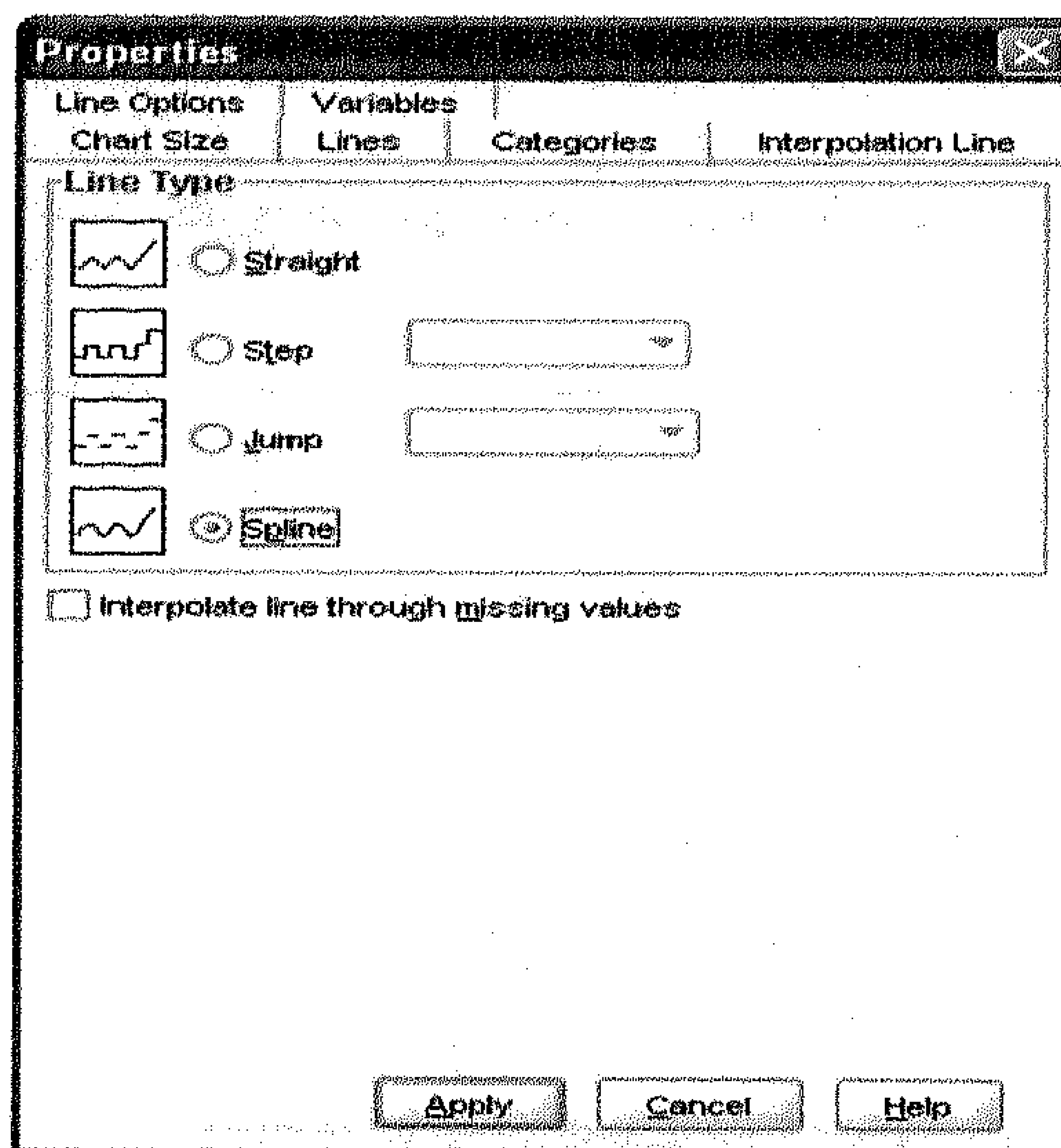
- تظهر نافذة جديدة ننقل المتغير x لخانة Category Axis



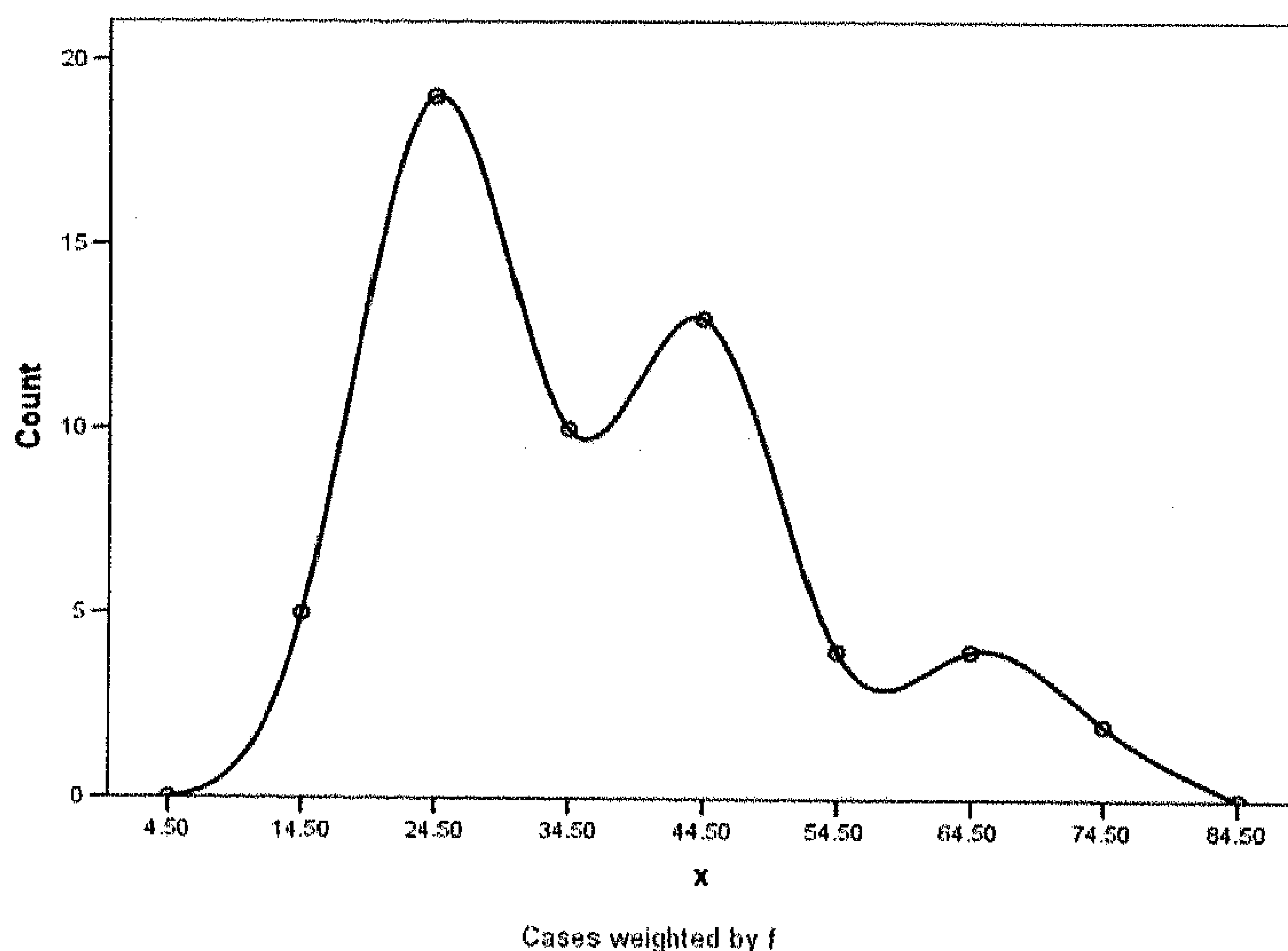
- نتأكد من اختيار N of cases من قائمة Line represents
- نضغط Ok فيظهر شكل المضلع التكراري التالي:



ولرسم المنحنى التكراري يمكن من نافذة المخرجات Output Viewer بالضغط على الشكل البياني مرتين متتاليتين ننتقل لمحرر الأشكال Chart Editor ثم بالضغط مرتين على الشكل البياني ننتقل لنافذة تنسيق الشكل البياني



وباختيار Interpolation Line يظهر نوع الخطوط التي تصل بين النقاط نختار منها Spline فيتحول الخط الواصل بين النقاط لمنحنى فنحصل على المنحنى التكراري كما بالشكل التالي:



٢-٤ التوزيعات التكرارية المتجمعة (Cumulative Frequency Distributions)

سوف نهتم هنا بكيفية تعيين التكرار المتجمع الصاعد وكيفية تمثيله بيانياً.

تطبيق (٢-٥):

في مثال (٢-١٠) أوجد التوزيع التكراري المتجمع الصاعد والتوزيع التكراري المتجمع الهابط للجدول التكراري الآتي:

طول النبات بالسنتيمتر	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50
عدد النباتات	12	16	30	25	12	5

الحل

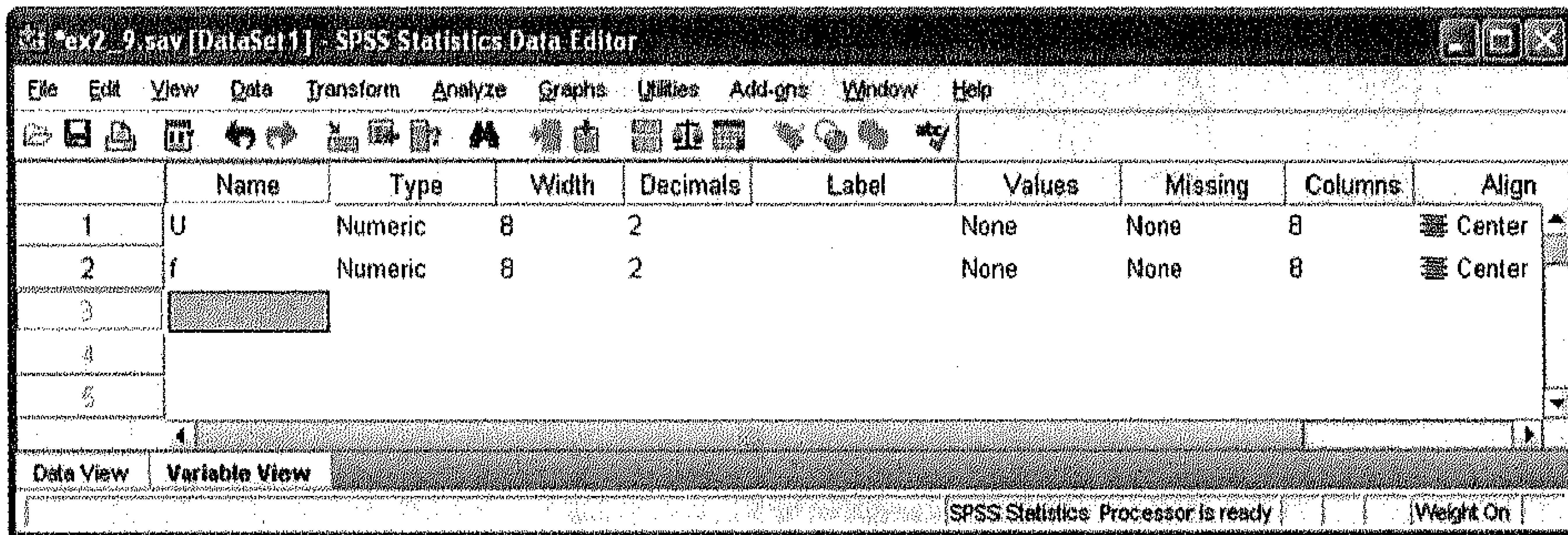
١- التوزيع التكراري المتجمع الصاعد

لتعيين (حساب) التكرار المتجمع الصاعد نقوم بالخطوات التمهيدية التالية:

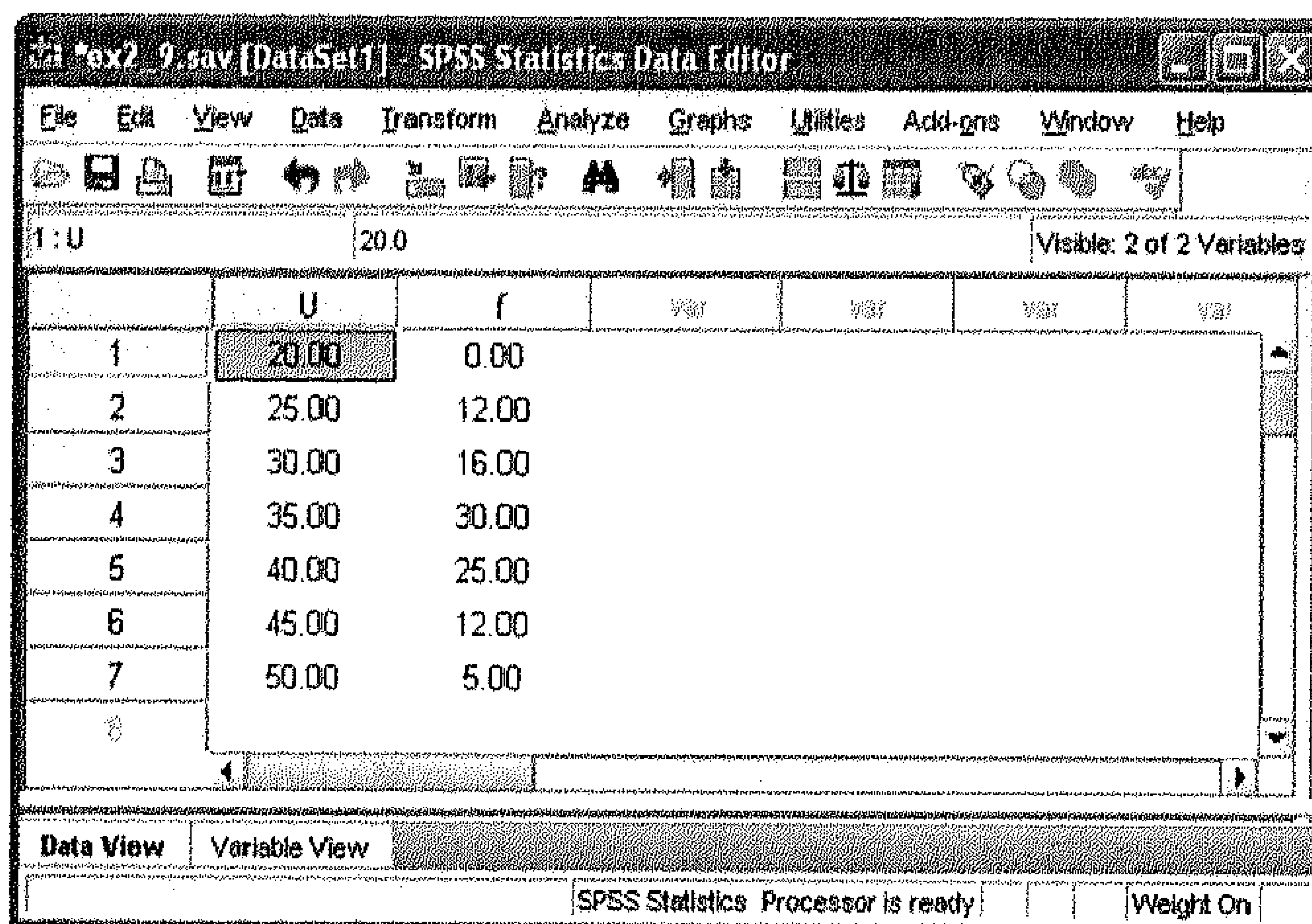
١- سوف نضيف فترة قبل أول فئة يقابلها تكرار صفر

الفترات	التكرار
15 – 20	0
20 – 25	12
25 – 30	16
30 – 35	30
35 – 40	25
40 – 45	12
45– 50	5

٢- نقوم بتعريف متغيرين أحدهما يمثل الحدود العليا للفترات U والثاني يمثل التكرار f في نافذة Variable view



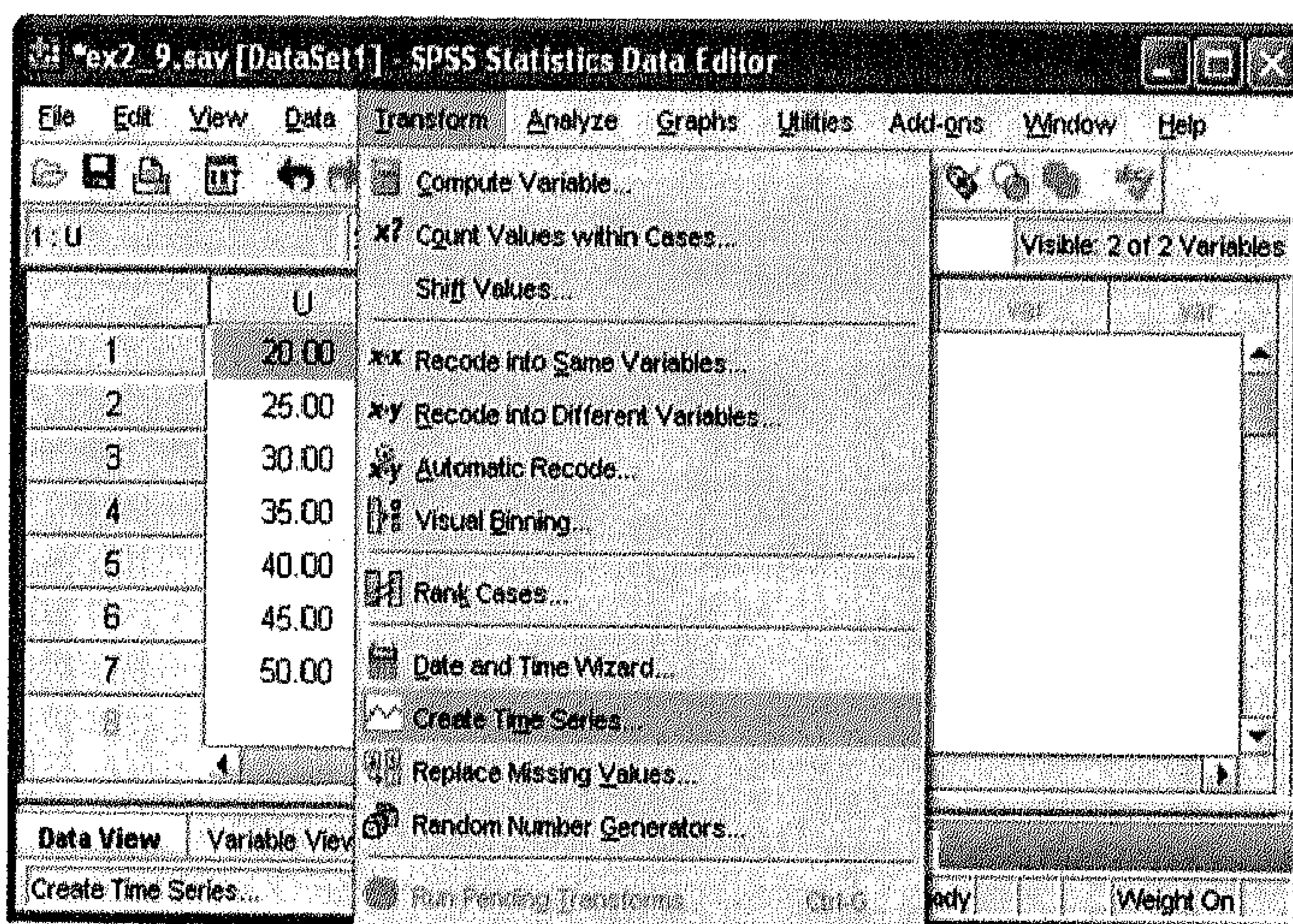
٣- ننتقل لنافذة Data view ونقوم بإدخال البيانات



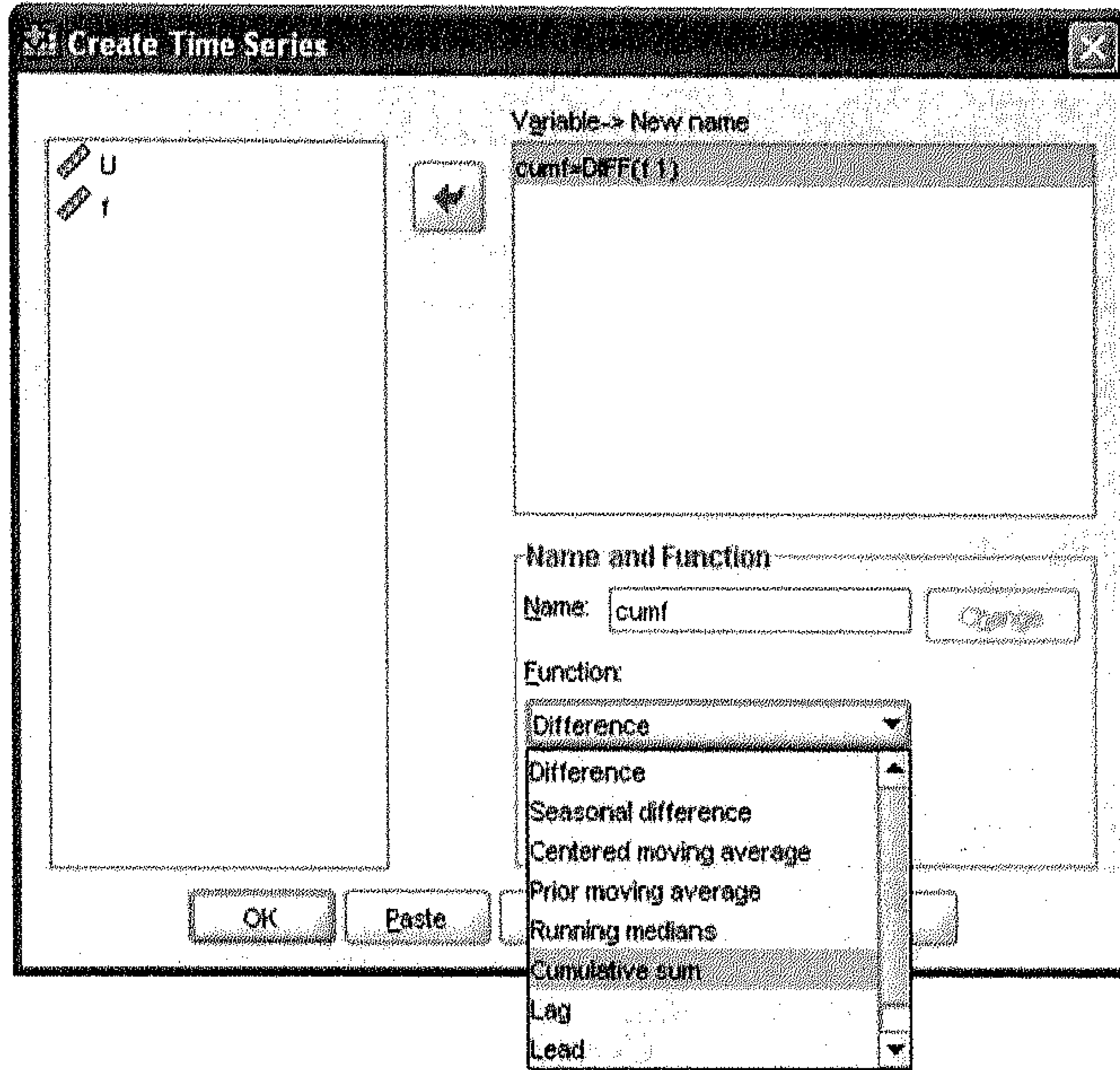
	U	f
1	20.00	0.00
2	25.00	12.00
3	30.00	16.00
4	35.00	30.00
5	40.00	25.00
6	45.00	12.00
7	50.00	5.00

٤- الآن يمكن تعيين أو حساب التكرار المتجمع الصاعد باتباع الخطوات التالية:

- من قائمة Transform نختار Create Time Series



- تظهر شاشة بعنوان Create Time Series سوف ننقل المتغير *f* لقائمة New variable(s)، من Name and Function نحدد اسم المتغير الجديد وليكن Cumf من قائمة Function نختار الأمر Cumulative Sum ثم نضغط على Change المقابل لشريط Name



- بالضغط على Ok والعودة إلى ملف البيانات نجد أنه قام بتعيين متغير جديد بعنوان Cumf يحتوى على التكرار المتجمع الصاعد

*ex2_9.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : U 20.0 Visible: 3 of 3 Variables

	U	f	cumf	var	var
1	20.00	0.00	0.00		
2	25.00	12.00	12.00		
3	30.00	16.00	28.00		
4	35.00	30.00	58.00		
5	40.00	25.00	83.00		
6	45.00	12.00	95.00		
7	50.00	5.00	100.00		
8					

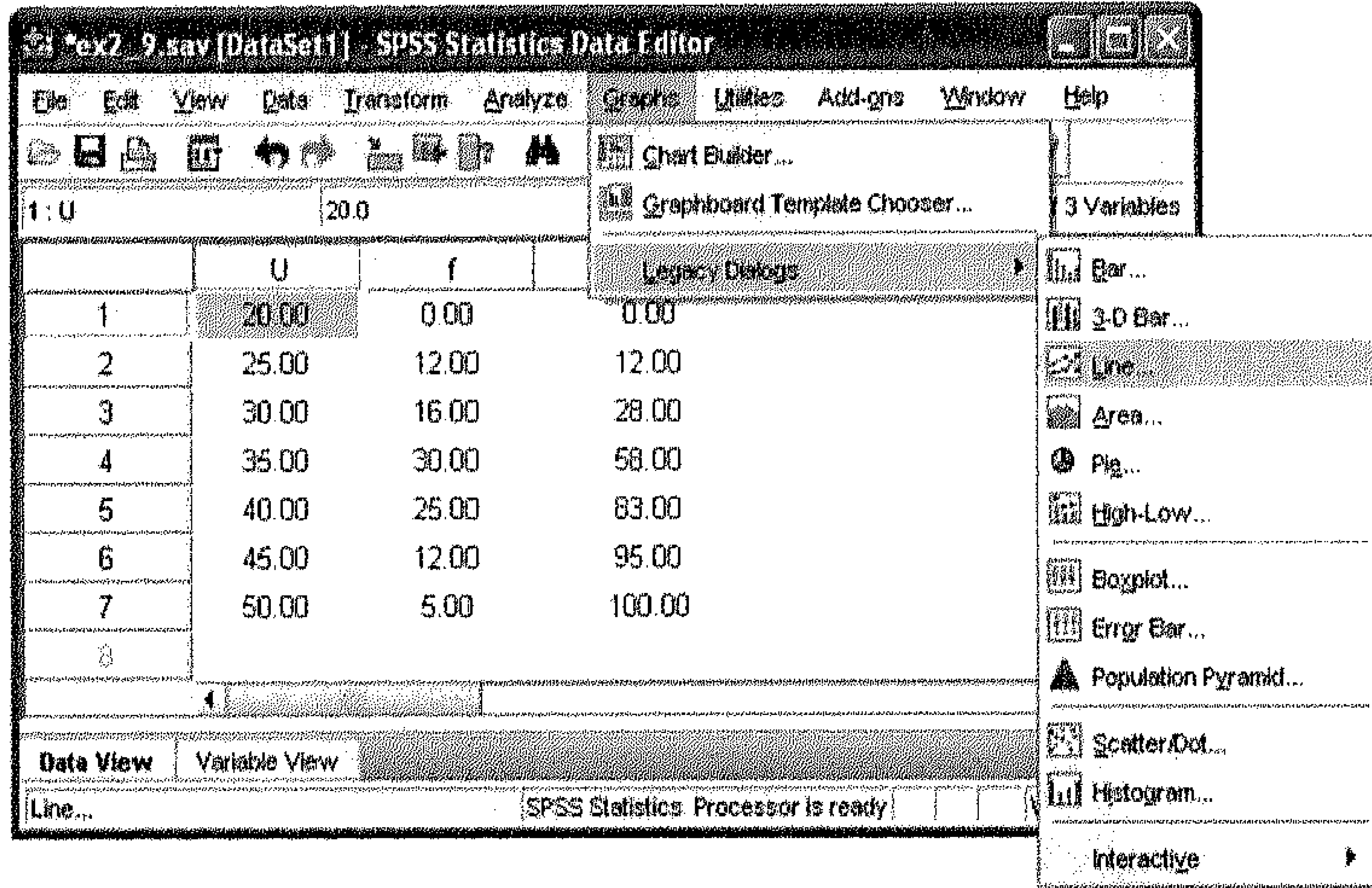
Data View Variable View

SPSS Statistics Processor is ready Weight On

٥- يُمكن رسم منحنى التكرار المتجمع الصاعد بطريقتين:

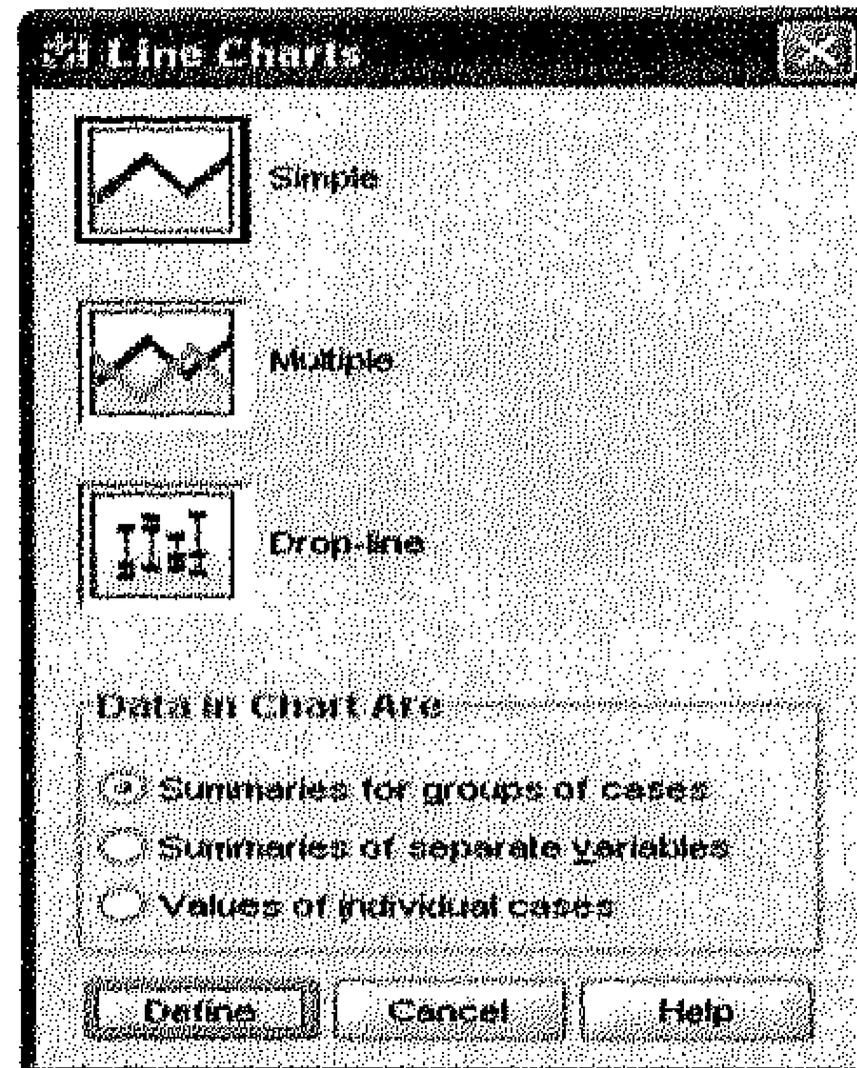
الطريقة الأولى: باستخدام التكرار المتجمع الصاعد Cumf والحدود العليا للفترات

- من قائمة Graphs نختار الأمر Line من القائمة Legacy Dialogs

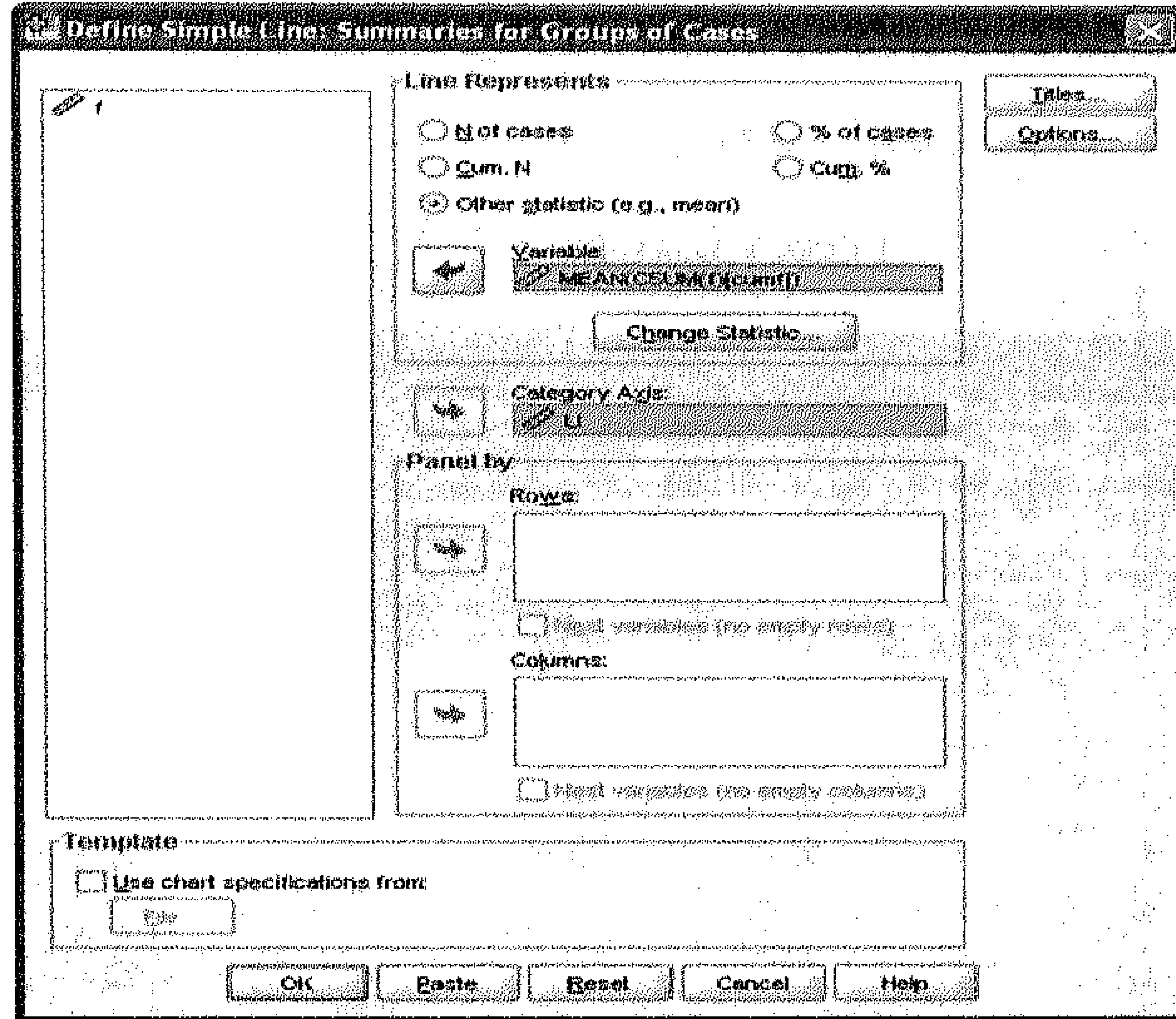


	U	f	
1	20.00	0.00	0.00
2	25.00	12.00	12.00
3	30.00	16.00	28.00
4	35.00	30.00	58.00
5	40.00	25.00	83.00
6	45.00	12.00	95.00
7	50.00	5.00	100.00

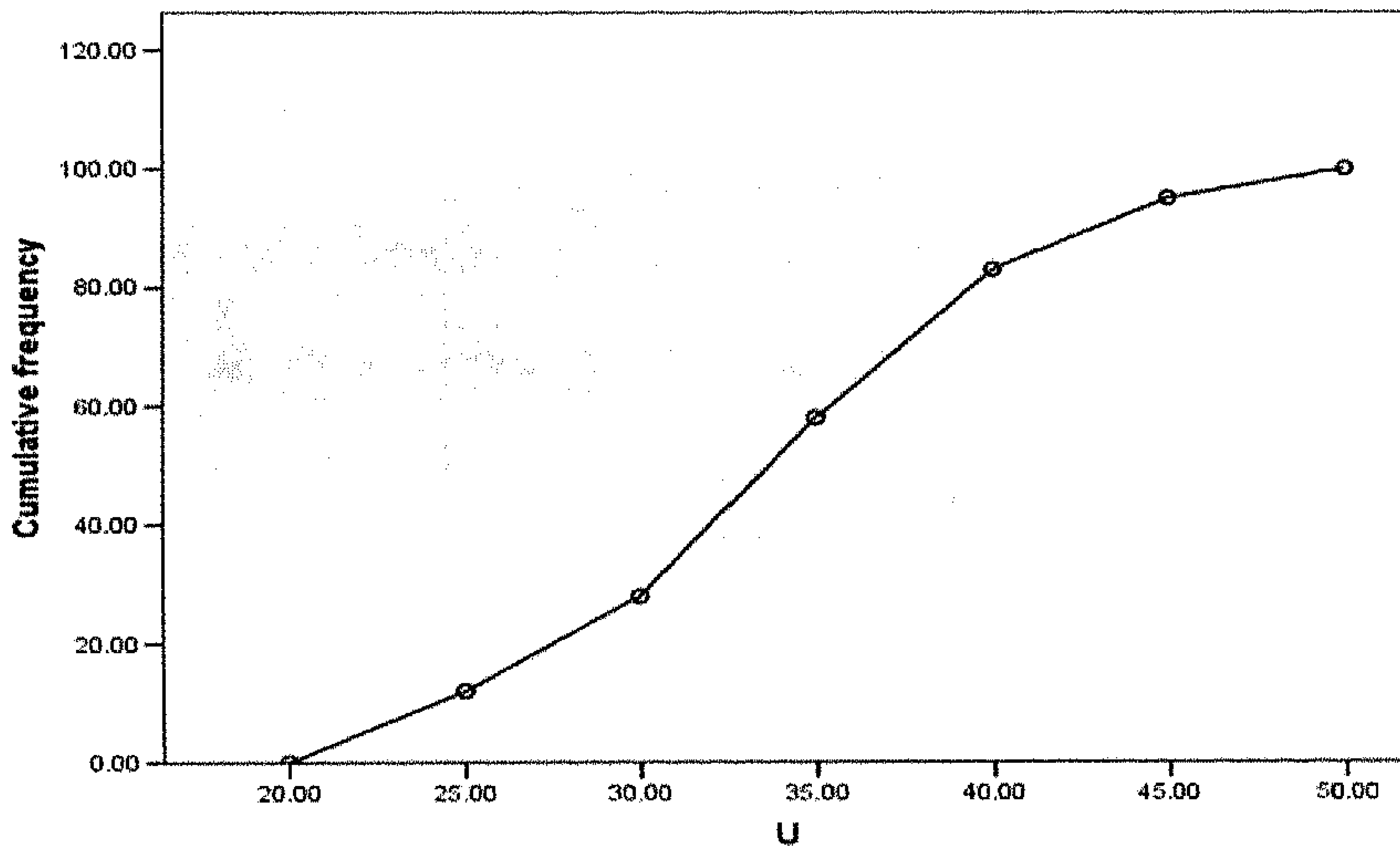
- تظهر نافذة جديدة بعنوان Line Charts



- نختار منها Simple ثم نضغط على Define
- ننقل المتغير U لخانة Category Axis ثم من قائمة Line Represents نختار Other statistics ثم ننقل المتغير Cumf لخانة Variable

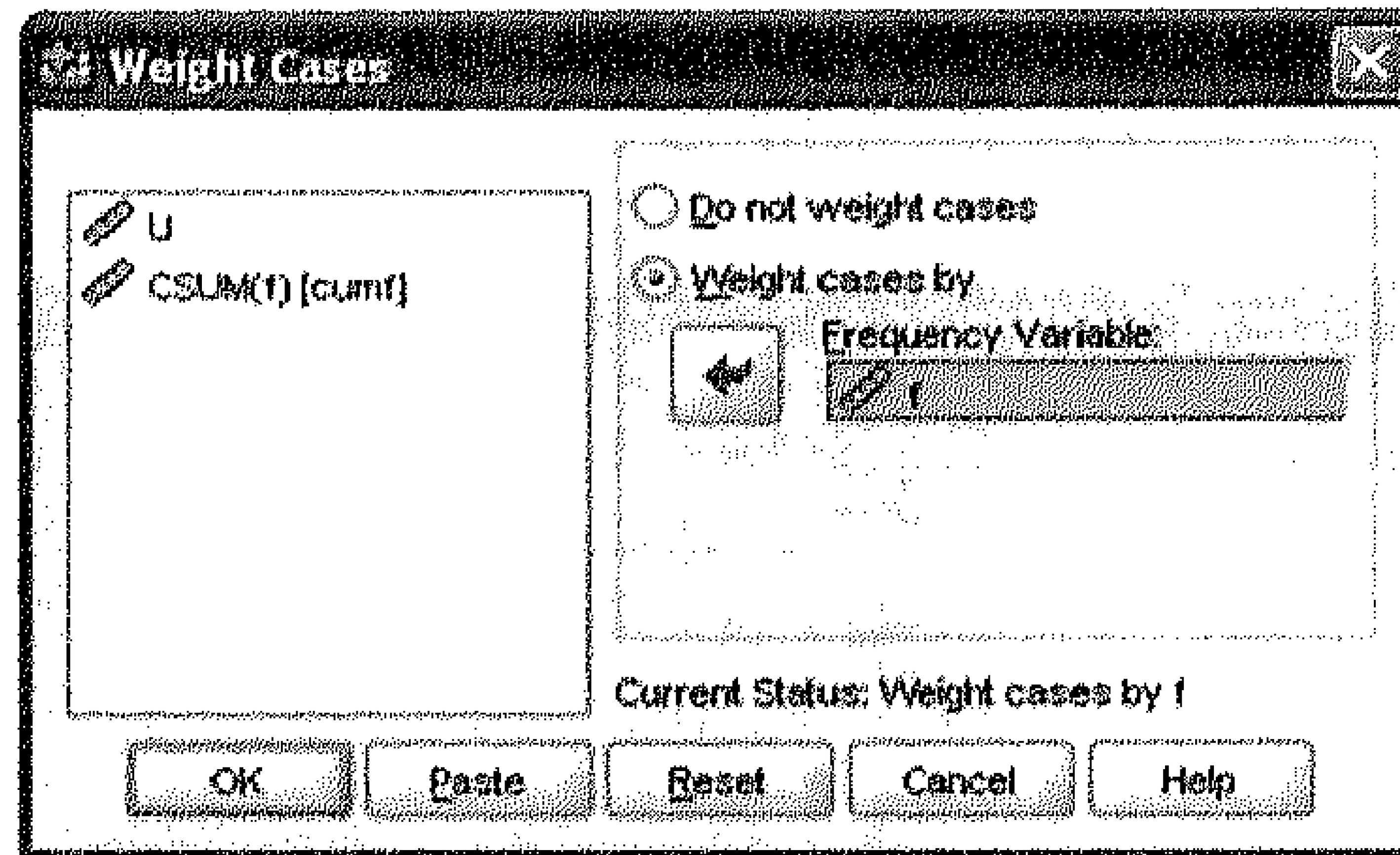


- بالضغط على Ok نحصل على الرسم التالي

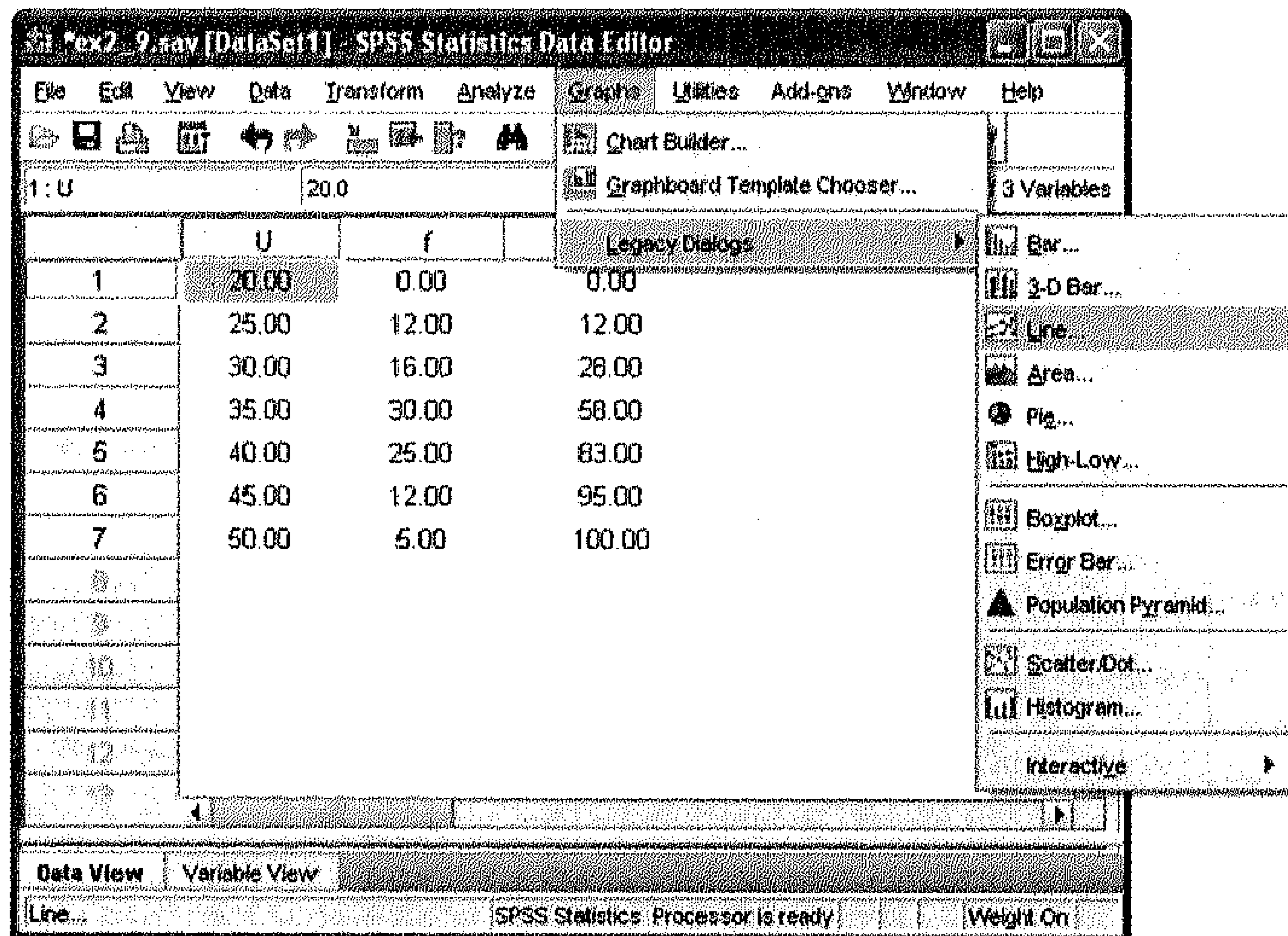


الطريقة الثانية: باستخدام التكرار f والحدود العليا للفترات

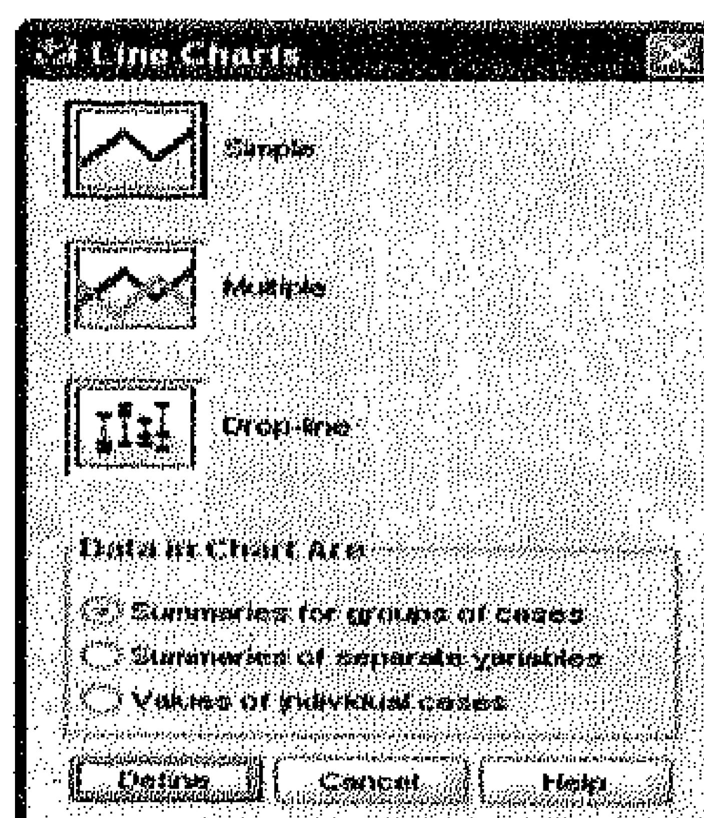
قبل الرسم يجب تحديد أن f هي تكرار للفترات التي حدها العلوي هو U وذلك باستخدام الأمر Weight cases من قائمة Data



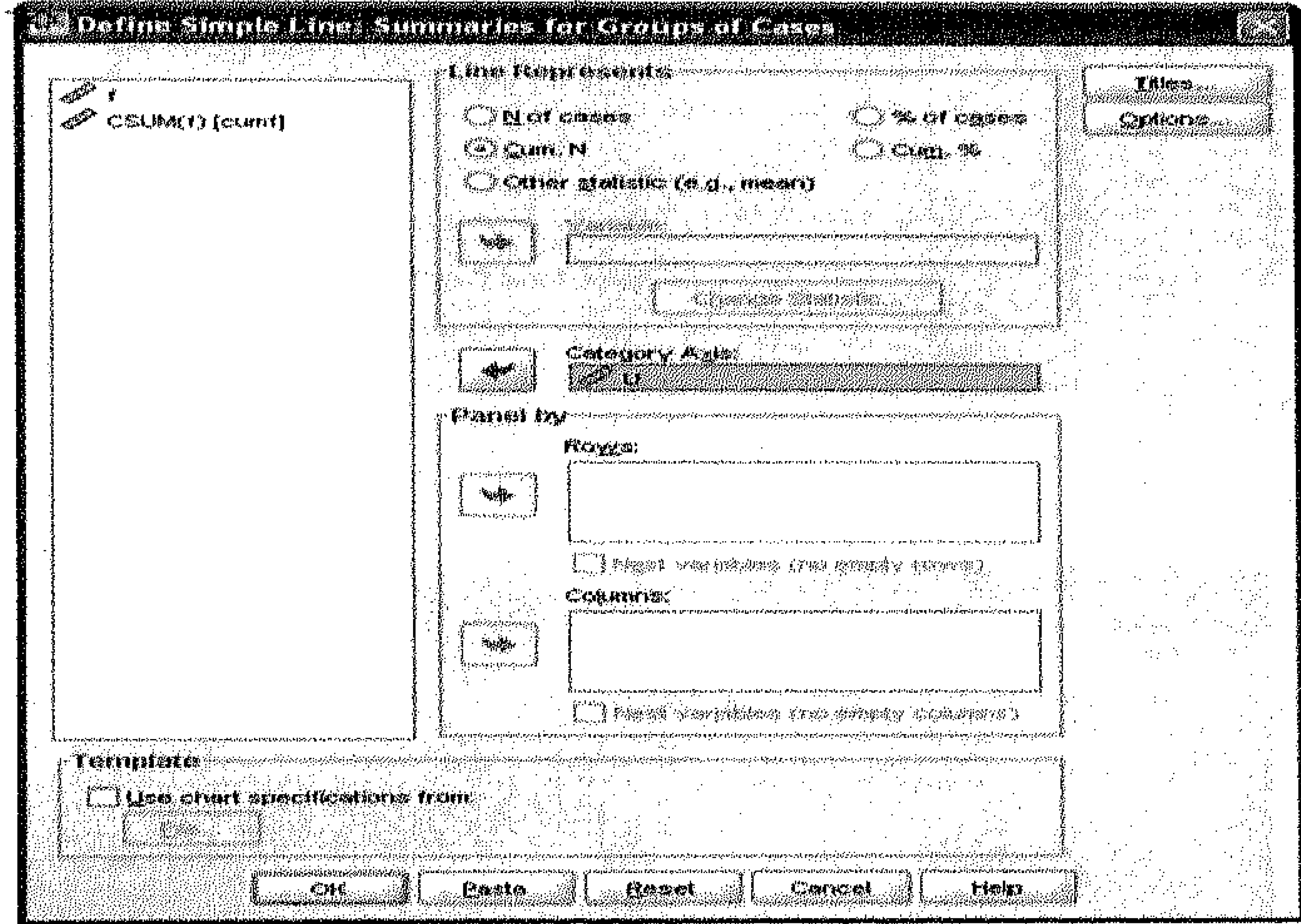
- من قائمة Graphs نختار الأمر Line من قائمة Legacy Dialogs



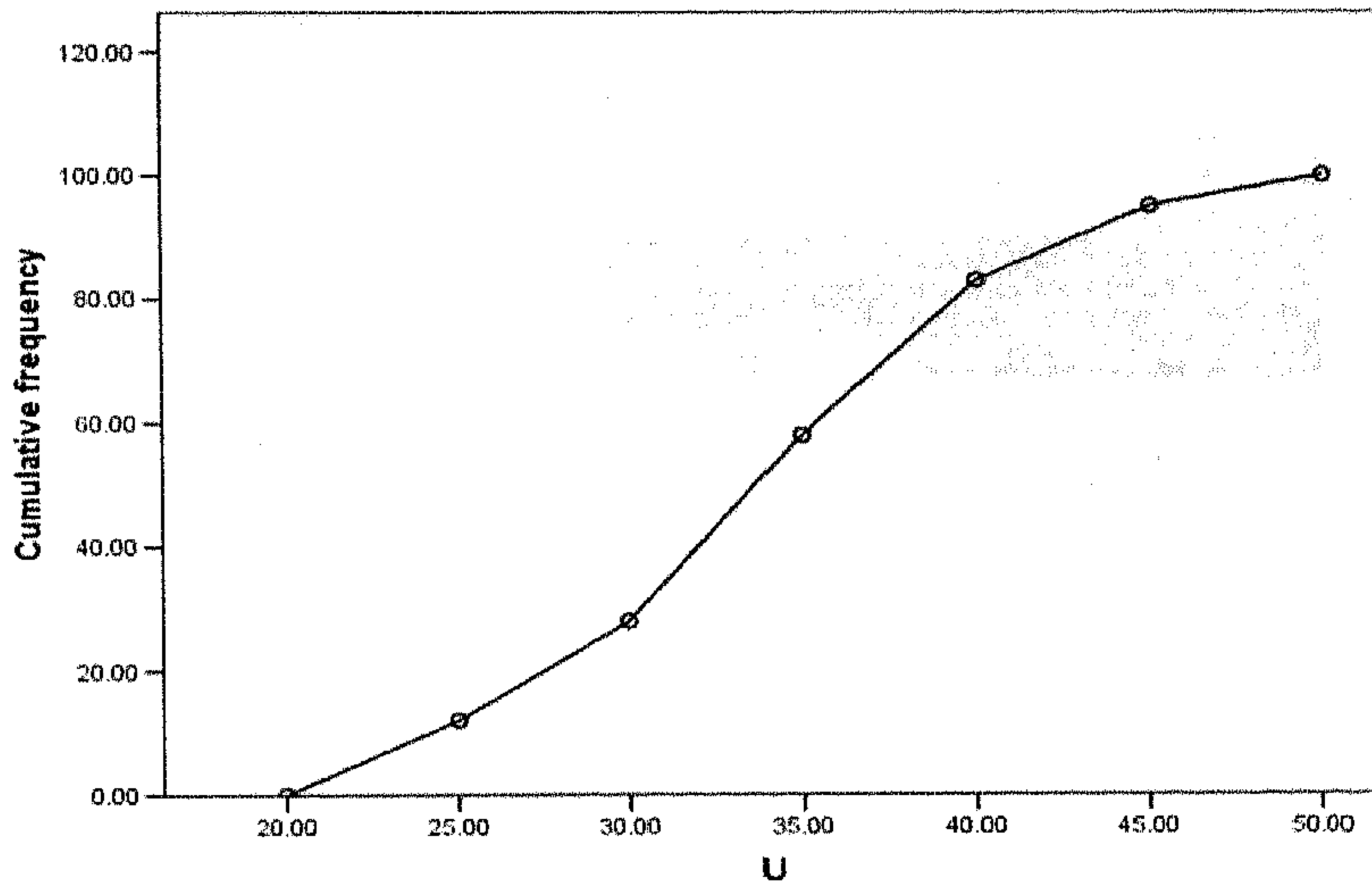
- تظهر نافذة بعنوان Line Charts نختار منها Simple ثم نضغط على Define



- ننقل المتغير U لخانة Category Axis ثم من قائمة Line Represents نختار Cum. N يمكن رسم التكرار المئوي المتجمع الصاعد باختيار Cum % من قائمة Line Represents



- ثم نضغط على Ok فنحصل على الرسم التالي:



أسئلة وتمارين (٢)

١- أشرح معنى كل من: التكرار، التكرار النسبي، التكرار المتجمع، مركز الفئة، طول الفئة.

٢- ما المنحنى التكراري؟ وما أنواعه؟

٣- كيف يمكن تمثيل البيانات الكمية المستمرة؟

٤- ما الفرق بين التكرار المتجمع الصاعد، التكرار المتجمع الهابط؟ ومتى يستخدم كل منهما؟

٥- إذا أردنا تعيين عدد العناصر التي تقع في فترة معينه فإيهما أفضل في الاستخدام منحنى التكرار المتجمع الصاعد أم منحنى التكرار المتجمع الهابط؟

٦- ما الفرق بين الأعمدة البيانية والدائرة البيانية وأيهما أفضل عند مقارنة قيم أكثر من متغير؟

٧- متى يستخدم الجدول التكراري ذو الفترات لجدولة البيانات؟

٨- ما هو المقصود بالتكرار المعدل؟ ومتى يستخدم؟

٩- اختر الإجابة الصحيحة من بين الإجابات التابعة للعبارات التالية:

- مجموع التكرار النسبي يساوي

(i) 1 (ii) 100 (iii) n (iv) ليس ما سبق

- يتم تمثيل الجدول التكراري للبيانات الوصفية باستخدام

(i) الدائرة البيانية (ii) المدرج التكراري (iii) المضلع التكراري (iv) كل ما سبق

- يستخدم في تعيين التوزيع التكراري المتجمع الصاعد

(i) مركز الفترات (ii) الحدود العليا للفترات (iii) الحدود الدنيا للفترات (iv) كل ما سبق

١٠-اكتب المصطلح المناظر للعبارات التالية:

- ناتج قسمة التكرار المناظر للفئة على مجموع التكرارات
- مجموع التكرارات ابتداء من أول تكرار حتى التكرار المناظر لفئة معينة
- الفرق بين التكرار المتجمع الصاعد المناظر للحد U_1 والحد U_2
- الفرق بين الحد الأدنى لفئة ما والحد الأدنى للفئة التالية لها.
- نصف مجموع حدي الفترة

- عدد القيم التي تقع داخل حدود فترة ما.

١١-فيما يلي بيانات عن لون العينين لعينة تم اختيارهم عشوائياً من أحد المجتمعات:

أسود	أسود	أزرق	أسود	بني	أزرق	أخضر	أسود	بني
أسود	أزرق	أسود	بني	أزرق	أخضر	أسود	بني	أخضر
أخضر	أخضر	بني	أزرق	أسود	أسود	أزرق	أخضر	اسود
بني	أزرق	أزرق	أخضر	بني	أسود	بني	أسود	أزرق

كون جدول توزيع تكراري لهذه البيانات ثم اعرضه بيانياً.

١٢- في 40 أسرة من سكان مدينة الرياض تم تحديد عدد الأولاد في كل أسرة وكانت البيانات كالتالي:

2	4	0	3	2	1	5	4	0	1
1	3	4	0	3	1	2	5	4	0
5	2	3	4	0	3	4	2	1	5
2	0	5	2	3	4	0	1	2	3

كون جدول توزيع تكراري لهذه البيانات ثم اعرضه بيانياً.

١٣- إذا كانت (X) ترمز إلى الوقت (بالساعة) الذي يستغرقه إصلاح آلة متعطلة في أحد المصانع. ومن سجلات هذا المصنع حصلنا على هذه البيانات للمتغير (X)

43.2	26.2	61.2	12.3	51.3	54.0	17.4	11.1	53.7
15.5	52.8	32.7	60.3	52.8	42.1	57.9	64.8	62.8
37.9	58.0	39.5	41.7	50.8	67.9	21.0	25.6	49.3
63.6	44.0	29.2	65.3	55.4	33.6	68.7	46.4	58.5
42.3	23.1	66.3	61.2	62.1	56.9	31.4	63.1	20.7

(أ) كون جدول التوزيع التكراري للمتغير X ومثله بيانياً

(ب) أوجد عدد الآلات التي يستغرق وقت إصلاحها وقتاً أقل من 30 ساعة.

(ج) أوجد عدد الآلات التي يبلغ وقت إصلاحها 60 ساعة أو أكثر.

١٤- فيما يلي التوزيع التكراري لأعضاء هيئة التدريس بإحدى الجامعات السعودية حسب جنسيتهم

الجنسية	سعودي	مصري	سوري	أردني	هندي	باكستاني	جنسيات أخرى
العدد	500	110	70	85	200	180	300

(أ) أوجد التكرار المئوي والنسبي لجنسية أعضاء هيئة التدريس.

(ب) ارسم التوزيع التكراري لجنسية أعضاء هيئة التدريس.

١٥- فيما يلي التوزيع التكراري لمائة قطعة أرض زراعية متساوية المساحة حسب إنتاجها من القمح (الكيلوجرام) في أحد الفصول الزراعية

الإنتاج	80 – 89	90 – 99	100 – 109	110 – 119	120 – 129	130 – 139
عدد القطع	8	10	14	21	28	9

(أ) ارسم كلاً من المدرج التكراري والمضلع التكراري.

(ب) ما نسبة قطع الأراضي التي يقل إنتاجها عن 105 كيلوجراماً؟

(ج) إذا اعتبر أن 15% من القطع إنتاجها رديء فما هو حجم الإنتاج اللازم لكي تعتبر قطعة الأرض ليست رديئة من حيث الإنتاج؟

١٦- البيانات التالية عبارة عن أوزان 50 عبوة من عبوات الحلوى بالكيلوجرام والتي تم بيعها خلال أسبوع

2.5	4.0	1.5	2.5	3.0	4.1	2.5	2.4	2.0	1.4
2.5	1.6	1.3	2.4	3.1	2.2	3.5	2.0	1.5	1.3
1.2	3.4	2.6	3.2	3.3	3.1	2.9	1.9	2.2	2.1
2.4	1.7	1.1	1.9	2.2	2.6	1.8	1.2	3.5	3.4

(أ) كون الجدول التكراري لتلك البيانات ومثله بيانياً.

(ب) عين التوزيع التكراري النسبي والتكراري المئوي.

(ج) عين التوزيعات التكرارية المتجمعة ومثلها بيانياً.

١٧- يجرى أحد الباحثين تجربه لقياس الزيادة في الوزن بالجرام بعد إعطاء عقار معين لمجموعة من الفئران وسجلت النتائج التالية:

الزيادة في الوزن	0 – 9.9	10 – 14.9	15 – 19.9	20 – 24.9
التكرار	25	30	20	15

(أ) أوجد التوزيع التكراري النسبي والمئوي.

(ب) ارسم المدرج التكراري والمضلع التكراري.

(ج) أوجد التوزيع التكراري المتجمع الصاعد والهابط.

(د) أوجد نسبة الفئران التي يتراوح مقدار الزيادة في وزنها بين 5 إلى 13 جراماً.

(هـ) أوجد عدد الفئران التي مقدار الزيادة في وزنها أكثر من 16 جراماً.

١٨- البيانات التالية تمثل درجات الطلاب بالاختبار النهائي لمقرر 1060 إحص لطلاب البرنامج الموحد للكليات الصحية في الفصل الدراسي الأول في أحد الأعوام الدراسية السابقة

69	50	74	71	81	71	75	68	71
66	73	53	56	63	70	72	57	53
80	67	85	72	75	68	51	42	79
93	91	72	61	66	57	83	60	77
76	50	57	60	55	53	73	82	60
66	72	72	65	67	64	64	78	87

(أ) أوجد التوزيع التكراري ومثله بيانياً.

(ب) أوجد التوزيع التكراري المتجمع الصاعد.

(ج) أوجد عدد الطلاب الحاصلين على درجات أكبر من 80 درجة.

(د) أوجد نسبة الطلاب الراسبين.

١٩- البيانات التالية تمثل الأجر الأسبوعي بالدولار الأمريكي لخمسين موظفاً في إحدى الشركات:

57	43	46	24	44	38	19	54	49	57
29	53	45	47	41	37	49	56	47	29
31	32	51	52	42	45	28	43	49	34
24	42	28	39	21	18	37	34	29	23
35	43	39	41	26	27	32	28	33	37

(أ) أوجد التوزيع التكراري للأجر الأسبوعي للموظفين بتلك الشركة ومثله بيانياً.

(ب) عين عدد الموظفين الذين يقل أجرهم عن 20 دولاراً.

(ج) أوجد نسبة الموظفين الذين يتراوح أجرهم بين 40، 50 دولاراً.

٢٠- البيانات التالية تمثل أعمار مجموعة من الموظفين بإحدى المؤسسات الخاصة بالمملكة العربية السعودية

فئات العمر	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54	المجموع
عدد الموظفين	5	10	20	5	40	100

(أ) أوجد التكرار النسبي والمئوي لهذه البيانات.

(ب) ارسم المدرج التكراري، المنحنى التكراري.

(ج) أوجد التوزيع التكراري المتجمع الهابط لهذه البيانات ومثله بيانياً.

٢١- ارسم كلاً من المدرج التكراري والمدرج التكراري المتجمع الصاعد للبيانات التالية:

الفترات	150 –	155 –	165 –	185 – 195
التكرار	15	50	40	30

٢٢- البيانات التالية تمثل أعمار مجموعة من سكان مدينة الرياض

الفترات	40 –	60 –	80 –	100 –	120 – 140
التكرار	9	12	20	6	1

(أ) أوجد عدد السكان الذين تقل أعمارهم عن 80 سنة.

(ب) عين عدد السكان الذين تقل أعمارهم عن 55 سنة.

(ج) ما نسبة السكان الذين تزيد أعمارهم عن 110 سنة؟

٢٣- البيانات في الجدول التالي تبين أعداد المواليد والوفيات في إحدى المدن خلال السنوات 1980/1984م

السنة	المواليد	الوفيات
1980	50	40
1981	70	73
1982	40	50
1983	60	58
1984	70	76

مثّل هذه البيانات بيانياً ثم قارن بينها.

٢٤- لقد قام أحد الباحثين بتجميع بيانات فصائل الدم التي قام أحد بنوك الدم باختبارها خلال أسبوع في المستشفى الجامعي وكانت البيانات كالتالي:

A	A	B	O	A	B	AB	O	B	A
O	AB	O	O	A	A	B	B	B	O
A	B	A	A	O	B	AB	O	AB	A
AB	O	A	B	B	O	O	A	B	AB

(أ) أحسب كلاً من التكرار النسبي والتكرار المئوي لفصائل الدم.

(ب) مَثّل التوزيع التكراري باستخدام الدائرة البيانية.

٢٥- البيانات التالية تمثل علامات 30 طالباً في امتحان درجته النهائية من 20 درجة موزعة كالتالي:

العلامات	6 – 8	9 – 11	12 – 14	15 – 17	18 – 20
عدد الطلاب	0	3	4	5	8

(أ) مثل التوزيع التكراري باستخدام المضلع التكراري.

(ب) أحسب التكرار النسبي والتكرار المئوي لعلامات الطلاب.

(ج) إذا كان 20% من الطلاب تقل درجاتهم عن علامة معينة فما تلك العلامة؟

٢٦- البيانات التالية تمثل الأرباح السنوية بآلاف الريالات السعودية لـ 30 محلاً تجارياً في إحدى المدن موزعة كما يلي:

فئات الربح	30 – 34	25 – 29	20 – 24	15 – 19	10 – 14
التكرار	5	15	6	4	0

(أ) أحسب مراكز الفترات(الفئات).

(ب) أحسب التوزيعات التكرارية المتجمعة للأرباح السنوية، ومثلها بيانياً

٢٧- في الجدول التكراري التالي توزيع 500 موظفاً حسب الأجر الشهري بالريال السعودي، بناء على بيانات العينة المختارة من مجتمع العاملين في إحدى الشركات

الفترات	500 – 1000	250 –	100 –	0 –
عدد العمال	25	125	150	200

والمطلوب

(أ) إيجاد جدول التكرار المعدل.

(ب) رسم المضلع التكراري لجدول التكرار المعدل.

(ج) هل منحنى توزيع الأجور متماثل؟

(د) حساب نسبة العمال الذين تزيد أجورهم عن 75 ريالاً.

(هـ) إذا كان 10% من العمال تزيد أجورهم عن قيمة معينة، فما هي تلك القيمة؟

٢٨- البيانات التالية تمثل أوزان 48 طفلاً بالرطل في أحد مراكز الرعاية الصحية بالمملكة العربية السعودية

68	63	42	27	30	23	36	28	32	79	46	27
22	23	24	25	44	19	65	43	25	74	30	51
36	42	28	31	28	43	25	45	12	57	12	51
12	32	49	38	42	49	27	31	50	38	28	21

والمطلوب

(أ) تكوين الجدول التكراري لأوزان هؤلاء الأطفال،

(ب) التكرار المتجمع الصاعد لأوزان الأطفال.

الفصل الثالث

مقاييس النزعة المركزية

MEASURES OF CENTRAL TENDENCY

٣–١ مقدمة

كان هدفنا في الفصل الثاني هو التعرف على الطرق والأساليب التي تستخدم في تلخيص وتصنيف وتبويب البيانات الإحصائية وكذلك التعرف على الأشكال البيانية التي يمكن أن تعرض بها هذه البيانات، ويمكننا الآن الانتقال إلى الخطوة التالية وهى حساب بعض المقاييس التي تصف البيانات الكمية التي تم جمعها عن أحد المتغيرات موضع الدراسة. تنقسم المقاييس الإحصائية لنوعين من المقاييس؛ مقاييس الموضع وتُعرف بمقاييس النزعة المركزية ومقاييس تحدد شكل المنحنى للبيانات وتسمى مقاييس التشتت، وهدفنا في هذا الفصل هو التركيز على مقاييس النزعة المركزية وسوف نتعرض لمقاييس التشتت في الفصل التالي.

٣–٢ مقاييس النزعة المركزية (Measures of Central Tendency)

غالباً ما نلاحظ أن البيانات تميل إلى التمركز حول قيمة معينة، وفي هذه الحالة يمكن استخدام هذه القيمة المركزية (central value) لتمثيل هذه المجموعة من البيانات. والمقاييس المستخدمة في التعرف على هذه القيمة المركزية تسمى مقاييس النزعة المركزية (measures of central tendency) أو أحياناً المتوسطات (averages) ويجب أن تتوفر في مقياس النزعة المركزية الصفات الآتية لكي يكون مقياساً جيداً:

(أ) أن يعتمد المقياس في حسابه على كل المشاهدات.

(ب) أن يكون المقياس سهل الحساب والفهم.

(ج) أن تتوفر في المقياس القابلية للتعامل الجبري.

(د) أن لا يتأثر المقياس بوجود القيم المتطرفة والشاذة.

من بين مقاييس النزعة المركزية الوسط الحسابي، الوسيط، المنوال ولا تتوفر كل الصفات السابق ذكرها في مقياس واحد ولكن كل واحد من هذه المقاييس يفضل استخدامه في حالات ولا يفضل في حالات أخرى. وفى كل هذه المقاييس نفترض أن لدينا بيانات عن أحد المتغيرات (وليكن X) تم جمعها من مجموعة من المفردات عددها n تسمى هذه البيانات مشاهدات (observations) ويرمز لها برموز $x_1, x_2, \dots, x_n$ وتستخدم هذه المشاهدات في إيجاد مقاييس النزعة المركزية.

٣-٣ الوسط الحسابي (Arithmetic Mean)

يعتبر المتوسط(الوسط) الحسابي من أهم مقاييس النزعة المركزية على الإطلاق، وتختلف عملية حساب الوسط الحسابي حسب طبيعة البيانات هل هي بيانات لمتغير منفصل أو متغير متصل، وكذلك هي بيانات خام أم بيانات مبوبة في جدول تكراري. وسوف يرمز للوسط الحسابي بالرمز $\bar{x}$.

- **البيانات الخام (المفردة) (Raw Data)**

إذا كان لدينا n من القيم $x_1, x_2, \dots, x_n$ لمتغير X فإننا نحصل على الوسط الحسابي بقسمة مجموع هذه المشاهدات على عددها أي أن

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

حيث $\sum_{i=1}^{n} x_i$ ترمز لمجموع المشاهدات.

مثال(٣-١)

فيما يلي أوزان 12 طالب من طلاب المستوى الأول بإحدى كليات جامعة الخرج (بالكيلو جرام).

63 76 77 51 65 103 69 81 58 91 78 64

فأوجد الوسط الحسابي لأوزان هؤلاء الطلاب.

الحل:

بما أن البيانات هي بيانات مفردة فإننا سوف نحسب الوسط الحسابي من العلاقة

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$= \frac{64 + 78 + 91 + \cdots + 76 + 63}{12}$$

$$= \frac{876}{12} = 73$$

والناتج لا يمثل وزن طالب بعينه ولكنه يمثل وزن الطالب في هذه العينة بصفة عامة. وبعض الأوزان للطلاب ستكون أكبر من الوسط الحسابي وبعضها أقل منه ويمكن أن نجد أيضاً طلاباً أوزانهم يساوى الوسط الحسابي. والوسط الحسابي تقع قيمته بين أصغر وأكبر قيمة. والفرق بين المشاهدة والوسط الحسابي يسمى انحراف المشاهدة عن الوسط الحسابي وهذه الانحرافات هي:

-10 3 4 -22 -8 30 -4 8 -15 18 5 -9

ومجموع هذه الانحرافات يساوي صفراً.

- **البيانات المبوبة (Grouped Data)**

نَعني بالبيانات المبوبة البيانات التي تم تلخصيها في جدول تكراري وسوف يحسب الوسط الحسابي في هذه الحالة من العلاقة:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_k f_k}{n} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i$$

حيث n هي عدد المشاهدات ونحصل عليها من جمع التكرارات معا.

$$n = \sum_{i=1}^{k} f_i$$

لكن قيمة x_i تختلف، ففي حالة المتغيرات المنفصلة ترمز للقيمة لكن في حالة المتغيرات المتصلة التي تم وضعها في جدول تكراري ذو فئات فإنها تمثل مركز الفئة.

مثال (٣-٢)

فيما يلي عدد أفراد 20 أسرة تم اختيارها عشوائياً من أحد أحياء مدينة الرياض. احسب الوسط الحسابي لعدد أفراد الأسرة.

عدد أفراد الأسرة	2	3	4	5	6	7
عدد الأسر f	2	2	5	6	2	3

الحل:

حيث إن البيانات موجودة في جدول تكراري لمتغير كمي منفصل فإن

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i$$

سوف نقوم بإنشاء الجدول التالي لتسهيل الحسابات

عدد أفراد الأسرة (x_i)	عدد الأسر (f_i)	$x_i . f_i$
2	2	4
3	2	6
4	5	20
5	6	30
6	2	12
7	3	21
المجموع	**20**	**93**

ومن الجدول السابق نجد أن

$$\sum_{i=1}^{k} x_i f_i = 93, \qquad n = \sum_{i=1}^{k} f_i = 20$$

$$\therefore \ \bar{x} = \frac{93}{20} = 4.65$$

مثال (٣-٣)

أوجد الوسط الحسابي لعدد حبات الطماطم في النبات الواحد إذا توفرت لدينا البيانات الآتية:

عدد حبات الطماطم في النبات	2	3	4	5	6	7	8	9
عدد النباتات f	4	5	10	29	30	18	3	1

الحل

حيث إن البيانات موجودة في جدول تكراري لمتغير كمي منفصل فإن

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i$$

وسوف نقوم بإنشاء الجدول التالي لتسهيل الحسابات

عدد حبات الطماطم في النبات x_i	عدد النباتات f_i	$x_i . f_i$
2	4	8
3	5	15
4	10	40
5	29	145
6	30	180
7	18	126
8	3	24
9	1	9
المجموع	**100**	**547**

ومن الجدول السابق فإن

$$\sum_{i=1}^{k} x_i f_i = 547, \qquad n = \sum_{i=1}^{k} f_i = 100$$

$$\therefore \quad \bar{x} = \frac{547}{100} = 5.47$$

مثال (٣-٤)

أوجد الوسط الحسابي للتوزيع التكراري الآتي:

الفئات	35–	40–	45–	50–	55–	60–65
التكرارات	7	18	23	24	16	12

الحل

حيث إن البيانات موجودة في جدول تكراري لمتغير كمي متصل فإن

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i$$

حيث x_i تمثل هنا مركز الفئات. وبإنشاء الجدول التالي لسهولة الحسابات

الفئات	التكرار f_i	مركز الفئات x_i	$x_i f_i$
35−	7	37.5	262.5
40−	18	42.5	765.0
45−	23	47.5	1092.5
50−	24	52.5	1260.0
55−	16	57.5	920.0
60−65	12	62.5	750.0
المجموع	**100**		**5050.0**

ومن الجدول السابق نجد أن

$$\sum_{i=1}^{k} x_i f_i = 5050, \qquad n = \sum_{i=1}^{k} f_i = 100$$

$$\therefore \ \bar{x} = \frac{5050}{100} = 50.5$$

بعض خصائص الوسط الحسابي:

١- مجموع انحرافات القيم عن الوسط الحسابي $\bar{x}$ يساوى الصفر.

٢- مجموع مربعات انحرافات القيم عن الوسط الحسابي $\bar{x}$ أصغر من أو يساوى مجموع مربعات انحرافات القيم عن أي قيمة حقيقية c.

٣- إذا كانت $x_1, x_2, \ldots x_n$ مجموعة من القيم ووسطها الحسابى $\bar{x}$ وكانت a, b مقدارين ثابتين فإن متوسط البيانات $ax_1 \pm b,\ ax_2 \pm b, \ldots, ax_n \pm b$ هو $a\bar{x} \pm b$

٤- وإذا كان لدينا k من المجموعات، وكان عدد المشاهدات في تلك المجموعات هو $n_1, n_2, \ldots, n_k$ ومتوسطاتها الحسابية هي $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ على التوالى فإن الوسط الحسابى للمجموعة الكلية الناتجة عن دمج تلك المجموعات معاً يعطى بالعلاقة التالية:

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + \cdots + n_k} = \frac{\sum_{i=1}^{k} n_i \bar{x}_i}{\sum_{i=1}^{k} n_i}$$

مثال (٣-٥):

إذا كان المستوى الثالث بكلية العلوم بجامعة الخرج يتكون من 60 طالباً. وإذا كان الوسط الحسابي لأوزان 25 طالباً من ذلك المستوى هو 61 كجم والوسط الحسابي لأوزان 35 طالباً آخر من نفس المستوى هو 58 كجم. أوجد الوسط الحسابي العام لجميع الطلاب.

الحل

من بيانات المثال نجد أن

بالنسبة للمجموعة الأولى فإن $n_1 = 25, \bar{x}_1 = 61$

بالنسبة للمجموعة الثانية فإن $n_2 = 35, \bar{x}_2 = 58$

فإن الوسط العام للمجموعتين معاً يعطى من العلاقة

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{25 \times 61 + 35 \times 58}{25 + 35}$$

$$= \frac{3555}{60} = 59.25$$

مميزات الوسط الحسابي:

يعتبر المتوسط من أفضل مقاييس النزعة المركزية ومن أكثرها استخداماً وذلك بما يتمتع به من مميزات جيدة ومنها:

(١) سهل التعريف والحساب ويخضع للعمليات الجبرية بسهوله.

(٢) يأخذ في الاعتبار جميع المشاهدات (البيانات).

(٣) وحيد لمجموعة البيانات الواحدة.

عيوب الوسط الحسابي:

بالرغم من أن المتوسط يعتبر من أفضل مقاييس النزعة المركزية إلا أن له بعض العيوب نذكر منها ما يلي:

(١) لا يمكن حسابه للبيانات الوصفية.

(٢) يتأثر بالقيم المتطرفة والشاذة.

ملاحظة (٣-١):

وحدة الوسط الحسابي هي نفس وحدة البيانات الأصلية. فإذا كانت وحدة البيانات هي الكيلوجرام فإن وحدة المتوسط هي الكيلوجرام.

٣-٤ المتوسط المرجح (الموزون) (Weighted Mean)

في بعض الأحيان يرافق المشاهدات ما يسمى بالأوزان، فعلى سبيل المثال إذا كان لدينا طالب يدرس في المستوى الأول بكلية العلوم ولديه خمس مقررات دراسية بعضها يُدرس ساعتين أسبوعياً والآخر يُدرس ثلاثة ساعات والبعض يدرس أربعة ساعات وأردنا حساب متوسط درجات الطالب في نهاية الفصل الدراسي (المستوى) فإنه يتم استخدام مقياس يسمى بالمتوسط الموزون أو المرجح والذي يرمز له بالرمز $\bar{x}_w$.

فإذا كان لدينا مجموعة من المشاهدات $x_1, x_2, \dots, x_n$ مقرونة بالأوزان $w_1, w_2, \dots, w_n$ فإن المتوسط المرجح لهذه المشاهدات يحسب من العلاقة:

$$\bar{x}_w = \frac{x_1 w_1 + x_2 w_2 + \cdots + x_n w_n}{w_1 + w_2 + \cdots + w_n} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

مثال (٣-٦)

أوجد متوسط الدرجات لأحد طلاب المستوى الثاني بكلية العلوم والدراسات الإنسانية بالخرج، إذا كانت درجاته كالتالي:

رمز المقرر	عدد ساعات المقرر	الدرجات
1080 إحص	3	85
1080 حيا	3	72
1080 فيز	3	77
1080 ريض	3	69
1080 نجم	2	79
123 نجم	3	82

الحل:

متوسط درجات الطالب سوف يحسب باستخدام المتوسط المرجح

$$\bar{x}_w = \frac{x_1w_1 + x_2w_2 + x_3w_3 + x_4w_4 + x_5w_5 + x_6w_6}{w_1 + w_2 + w_3 + w_4 + w_5 + w_6}$$

رمز المقرر	الدرجات x_i	عدد ساعات المقرر w_i	x_iw_i
1080 أحص	85	3	255
1080 حيا	72	3	216
1080 فيز	77	3	231
1080 ريض	69	3	207
1080 نجم	79	2	158
123 نجم	82	3	246
المجموع		17	1313

$$\bar{x}_w = \frac{1313}{17} = 77.235$$

ومن المعلوم أن الوسط الحسابي للبيانات ما هو إلا متوسط مرجح حيث إن

- الوسط الحسابي للبيانات المفردة هي متوسط مرجح لبيانات أوزان كل منها هو واحد.
- الوسط الحسابي للبيانات المبوبة هو متوسط مرجح حيث القيم هي مركز الفترات والأوزان هي التكرار المناظر لكل فترة.

٣–٥ الوسط الهندسي (The Geometric Mean)

الوسط الهندسي من مقاييس النزعة المركزية والذي يرمز له بالرمز G ، فالوسط الهندسي يمكن تعيينه لمجموعة من البيانات سواء كانت مفردة أو مبوبة.

- **البيانات المفردة (Raw Data)**

إذا كانت $x_1, x_2, \dots, x_n$ مجموعة من البيانات فإن الوسط الهندسي هو الجذر النوني لحاصل ضرب هذه الأرقام ويعطى بالعلاقة التالية:

$$G = \sqrt[n]{x_1x_2x_3 \dots x_n} = (x_1x_2x_3 \dots x_n)^{\frac{1}{n}}$$

مثال (٣-٧)

الوسط الهندسي للأرقام 2, 4, 8 هو

$$G = \sqrt[3]{2(4)(8)} = \sqrt[3]{64} = 4$$

ونجد أن قيمة الوسط الهندسي دائما أقل من قيمة الوسط الحسابي لمجموعة البيانات.

فالوسط الهندسي للكميتين 3, 12 هو

$$G = \sqrt[2]{x_1 x_2}$$

$$= \sqrt[2]{3 \times 12} = \sqrt[2]{36} = 6$$

لكن الوسط الحسابي للكميتين 3, 12 هو

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2}{2}$$

$$= \frac{3 + 12}{2} = \frac{15}{2} = 7.5$$

ويمكن أيضا تعيين المتوسط الهندسي باستخدام العلاقة التالية:

$$G = (x_1 x_2 x_3 \dots x_n)^{\frac{1}{n}}$$

وذلك بأخذ اللوغاريتم للأساس 10 للطرفين

$$\log_{10} G = \log_{10}(x_1 x_2 x_3 \dots x_n)^{\frac{1}{n}}$$

$$= \frac{1}{n}\log_{10}(x_1 x_2 x_3 \dots x_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \log_{10} x_i$$

وباستخدام اللوغاريتم العكسي للأساس 10 يمكن تعيين الوسط الهندسي. والوسط الهندسي يمتاز بعدم تأثره بالقيم المتطرفة ولكن لا يمكن استخدامه مع القيم السالبة أو الصفر لعدم جواز الجذر للقيم السالبة ذات العدد الفردي والقيمة صفر تلغي باقي القيم لكون الضرب في الصفر يساوي صفراً. ومن الملاحظ

أيضا أن $\log_{10} G$ هو المتوسط الحسابي للوغاريتم القيم $x_1, x_2, \dots, x_n$ مما يؤدى إلى أن قيمة المتوسط الهندسي أقل من قيمة المتوسط الحسابي.

مثال (٣-٨)

هل نقبل بالقيم السالبة في الوسط الهندسي؟ اكتب مجال للقيم التي يقبلها الوسط الهندسي.

الحل:

لا يقبل القيم السالبة لكونه جذر وإن وجدت قيمتان سالبتان (حاصل ضربهم موجب) المجال $]0, \infty[$

مثال (٣-٩)

أوجد الوسط الهندسي للقيم 2, 4, 4, 8

الحل:

يمكن حساب الوسط الهندسي باستخدام العلاقة التالية:

$$G = \sqrt[4]{x_1 x_2 x_3 x_4}$$

$$= \sqrt[4]{2 \times 4 \times 4 \times 8} = \sqrt[4]{256} = 4$$

يمكن حساب الوسط الهندسي باستخدام العلاقة التالية:

$$\begin{aligned}
\log_{10} G &= \frac{1}{n} \sum_{i=1}^{n} \log_{10} x_i \\
&= \frac{1}{n} [\log_{10}(x_1) + \log_{10}(x_2) + \cdots + \log_{10}(x_n)] \\
&= \frac{1}{4} [\log_{10}(2) + \log_{10}(4) + \log_{10}(4) + \log_{10}(8)] \\
&= \frac{1}{4} [0.30103 + 0.60206 + 0.60206 + 0.90309] \\
&= \frac{2.40824}{4} = 0.60206
\end{aligned}$$

وباستخدام معكوس اللوغاريتم للأساس 10 يمكن تعيين الوسط الهندسي G كالتالي:

$$\Longrightarrow G = (10)^{0.60206} = 4$$

البيانات المبوبة (Grouped Data):

يمكن تعيين الوسط الهندسي للبيانات المبوبة من العلاقة التالية

$$G = \sqrt[\Sigma f]{x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_k^{f_k}}$$

حيث f_i هو التكرار المقابل للقيمة x_i في حالة المتغير الكمي المنفصل، لكن في حالة المتغير الكمي المتصل هو التكرار المقابل للفترة التي مركزها x_i ويمكن كتابه العلاقة السابقة بدلالة اللوغاريتمات بالصورة التالية

$$G = \sqrt[\Sigma f_i]{x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_k^{f_k}}$$
$$= \left(x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_k^{f_k}\right)^{\frac{1}{\Sigma f_i}}$$

بأخذ اللوغاريتم للأساس 10 للطرفين

$$\log_{10}(G) = \frac{1}{\sum f_i} \log_{10}\left(x_1^{f_1} \times x_2^{f_2} \times \ldots \times x_k^{f_k}\right)$$
$$= \frac{1}{\sum f_i}[f_1 \log_{10}(x_1) + f_2 \log_{10}(x_2) + \cdots + f_k \log_{10}(x_k)]$$
$$= \frac{1}{\sum f_i} \sum_{i=1}^{k} f_i \log_{10}(x_i)$$

مثال(٣-١٠)

احسب الوسط الهندسي للبيانات المبينة في الجدول الآتي:

الفترات	20 – 39	40 – 59	60 – 79	80 – 99	100 – 119	120 – 139
التكرار	20	35	80	40	15	10

الحل:

- باستخدام قانون اللوغاريتم

$$\log_{10}(G) = \frac{1}{\sum f_i} \sum_{i=1}^{k} f_i \log_{10}(x_i)$$

الفترات	التكرار f_i	مركز الفترات x_i	$\log_{10}(x_i)$	$f_i \log_{10}(x_i)$
20 – 39	20	29.5	1.4698	29.3964
40 – 59	35	49.5	1.6946	59.3112
60 – 79	80	69.5	1.8420	147.3588
80 – 99	40	89.5	1.9518	78.0729
100 – 119	15	109.5	2.0394	30.5912
120 – 139	10	129.5	2.1123	21.1227
المجموع	200			365.8532

$$\therefore \log_{10}(G) = \frac{365.8532}{200} = 1.829266$$

$$\Longrightarrow G = (10)^{1.8293} = 67.49$$

باستخدام القانون الأول:

$$G = \sqrt[\Sigma f_i]{x_1^{f_1} \times x_2^{f_2} \times \dots \times x_k^{f_k}}$$

$$G = \sqrt[200]{(29.5)^{20} \times (49.5)^{35} \times (69.5)^{80} \times (89.5)^{40} \times (109.5)^{15} \times (129.5)^{10}}$$

$$= (29.5)^{\frac{20}{200}} \times (49.5)^{\frac{35}{200}} \times (69.5)^{\frac{80}{200}} \times (89.5)^{\frac{40}{200}} \times (109.5)^{\frac{15}{200}} \times (129.5)^{\frac{10}{200}}$$

$$= 1.4028 \times 1.9795 \times 5.4550 \times 2.4568 \times 1.4222 \times 1.2753$$

$$= 67.49$$

مميزات الوسط الهندسي:

(١) يتأثر بالقيم المتطرفة والشاذة.

(٢) سهل التعريف والحساب.

(٣) وحيد لمجموعة البيانات الواحدة.

عيوب الوسط الهندسي:

(١) لا يستخدم مع الصفر والقيم السالبة.

(٢) لا يمكن تعيينه للبيانات النوعية (الوصفية).

٣-٦ الوسط التوافقي (The Harmonic Mean)

ومن مقاييس النزعة المركزية التي لها تطبيقات عديدة الوسط التوافقي والذي يرمز له بالرمز H ويمكن تعيينه بسهولة كما حدث مع الوسط الحسابي والهندسي وأيضا الوسط المرجح

البيانات المفردة (Raw Data):

الوسط التوافقي لمجموعة البيانات $x_1, x_2, \dots, x_n$ يعطى من العلاقة التالية:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

حيث أن n ترمز لعدد البيانات(المشاهدات)، ومن الملاحظ أن $\frac{1}{H}$ تمثل المتوسط الحسابي للقيم $\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_n}$ وبذلك فإن قيمة المتوسط التوافقي أقل من قيمة المتوسط الحسابي.

فالوسط التوافقي للقيم 2, 4, 8 هو:

$$H = \frac{3}{\frac{1}{2} + \frac{1}{4} + \frac{1}{8}} = \frac{3}{\frac{7}{8}} = \frac{8(3)}{7} = 3.4286$$

لكن الوسط الحسابي يساوي

$$\bar{x} = \frac{2 + 4 + 8}{3} = \frac{14}{3} = 4.6667$$

البيانات المبوبة (Grouped Data):

في حالة البيانات المبوبة يمكن أيضاً تعيين الوسط التوافقي لتلك البيانات باستخدام العلاقة التالية:

$$H = \frac{f_1 + f_2 + \cdots + f_k}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_k}{x_k}} = \frac{\sum f_i}{\sum \left(\frac{f_i}{x_i}\right)}$$

مثال (٣-١١)

احسب الوسط التوافقي للبيانات المبينة في الجدول الآتي:

الفترات	20 – 39	40 – 59	60 – 79	80 – 99	100 – 119	120 – 139
التكرار	20	35	80	40	15	10

الحل:

الوسط التوافقي لتلك البيانات يعطى بالعلاقة التالية

$$H = \frac{\sum f_i}{\sum\left(\frac{f_i}{x_i}\right)}$$

الفترات	التكرار f_i	مركز الفترات x_i	f_i/x_i
20 – 39	20	29.5	0.6780
40 – 59	35	49.5	0.7071
60 – 79	80	69.5	1.1511
80 – 99	40	89.5	0.4469
100 – 119	15	109.5	0.1370
120 – 139	10	129.5	0.0772
المجموع	200		3.1973

$$\therefore H = \frac{200}{3.1973} = 62.5528$$

ملاحظة (٣-٢):

الوسط الحسابي ≤ الوسط الهندسي ≤ الوسط التوافقي أي أن:

$$H \leq G \leq \bar{x}$$

وتتحقق علامة التساوي إذا كانت الأرقام $x_1, x_2, x_3, \ldots, x_n$ متساوية. فمجموعة البيانات 8 ,4 ,2 وسطها الحسابي هو$\bar{x} = 4.667$ ووسطها الهندسى هو $G = 4$ ووسطها التوافقى هو $H = 3.429$

٣-٧ الوسيط (Median)

يعرف الوسيط على أنه القيمة التي يكون عدد المشاهدات الأكبر منها مساوياً لعدد المشاهدات الأقل منها بمعنى أنها القيمة التي تقسم البيانات لجزئيين متساوين أي يوجد قبلها 50% من البيانات وبعدها أيضا 50% من البيانات. ويرمز للوسيط بالرمز Med وتختلف طريقة تعيين الوسيط تبعاً لنوع البيانات.

- **البيانات المفردة (Raw Data):**

إذا كان لدينا القيم $x_1, x_2, \ldots, x_n$ فإنه لتعيين الوسيط لهذه القيم نقوم بالخطوات التالية:

١- نقوم بترتيب البيانات تصاعدياً أو تنازلياً.

٢- نحدد ما إذا كان عدد المشاهدات n فردياً أم زوجياً بحيث

- نحدد رتبة الوسيط فإذا كان عدد المشاهدات فردي فإنه توجد قيمة واحدة فقط في منتصف البيانات تُقسم البيانات لجزئيين متساوين ولذلك فإن الوسيط هو القيمة التي لها الترتيب $\frac{n+1}{2}$ وكما يلي فإن

$$Med = X_{\left(\frac{n+1}{2}\right)}$$

- أما إذا كان عدد المشاهدات زوجياً فإنه توجد قيمتين في المنتصف ولذلك فإن الوسيط هو الوسط الحسابي لهاتين القيمتين واللتين لها الترتيب $\frac{n}{2}, \frac{n}{2}+1$ وبالتالى فإن الوسيط هو

$$Med = \frac{1}{2}\left[X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right]$$

مثال (٣-١٢)

يريد صاحب أحد محلات الملابس الجاهزة تقدير وسيط العمر لعملائه من الإناث والذكور حتى يأخذ ذلك في الاعتبار عند شراء طلباته الجديدة. فأخذ عينة من عملائه الإناث وأخرى من الذكور فكانت أعمارهم بالسنوات هي:

الذكور (Y)	الإناث (X)
17 40 29 72 37	15 42 35 24 20 17 60 30 12 27

الحل

أولاً يجب ترتيب هذه القيم تصاعدياً لكي يمكن تعيين الوسيط

الترتيب	1	2	3	4	5	6	7	8	9	10
الإناث	12	15	17	20	24	27	30	35	42	60
الذكور	17	29	37	40	72					

- حيث عدد الذكور فردي $n_1=5$

فإنه يوجد قيمة واحدة فقط في المنتصف وهى القيمة التي لها الترتيب

$$\frac{n_1+1}{2} = \frac{5+1}{2} = \frac{6}{2} = 3$$

فإن الوسيط لأعمار الذكور هو القيمة

$$Med = X_{(3)} = 37$$

- وبالنسبة للإناث فإن عددها زوجي $n_2 = 10$

فإنه يوجد قيمتين في المنتصف واللتين لهما الترتيب

$$\frac{n_2}{2} = \frac{10}{2} = 5, \quad \frac{n_2}{2} + 1 = 5 + 1 = 6$$

فإن الوسيط لأعمار الإناث هو

$$Med = \frac{X_{(5)} + X_{(6)}}{2} = \frac{24 + 27}{2} = \frac{51}{2} = 25.5$$

- **البيانات المبوبة (Grouped Data):**

أولا المتغير المنفصل

في حالة المتغير الكمي المنفصل سيتم تعيين الوسيط عن طريق إتباع الخطوات التالية:

(أ) نقوم بتعيين التكرار المتجمع الصاعد.

(ب)نبحث نوع عدد البيانات n فردى أو زوجي.

- إذا كان عدد المشاهدات فردياً فإن الوسيط هو القيمة المقابلة للتكرار المتجمع الصاعد المساوي للمقدار $\frac{n+1}{2}$.
- إذا كان عدد المشاهدات زوجياً فإن الوسيط هي الوسط الحسابي للقيمتين المناظرتين للقيم $\frac{n}{2}, \frac{n}{2} + 1$ في التكرار المتجمع الصاعد.

مثال (٣-١٣)

فيما يلي التوزيع التكراري لعدد الحوادث اليومية خلال 100 يوم متتالية على احد الطرق:

X عدد الحوادث اليومية	0	1	2	3	4	5
f عدد الأيام	48	32	17	2	1	0

أوجد وسيط عدد الحوادث اليومية.

الحل

نلاحظ أن X متغير كمي متقطع (منفصل) وأن إجمالي عدد الأيام 100 = n وهو عدد زوجي فإننا يجب أن نحسب التكرار المتجمع الصاعد ونبحث فيه عن القيم

$$\frac{n}{2} = \frac{100}{2} = 50, \qquad \frac{n}{2} + 1 = 51$$

والتكرار المتجمع الصاعد هو

عدد الحوادث اليومية x_i	عدد الأيام f_i	التكرار المتجمع الصاعد
0	48	48
1	**32**	**80**
2	17	97
3	2	99
4	1	100
5	0	100

نجد أنه لا توجد أي من القيم 51 ,50 ضمن قيم التكرار المتجمع الصاعد لذا سوف نختار القيم التالية لها والموجودة وهى القيمة 80 لذا فإن:

$$X_{(50)} = 1, \; X_{(51)} = 1$$

وبذلك يكون الوسيط لعدد الحوادث اليومية هو:

$$Med = \frac{X_{(50)} + X_{(51)}}{2} = \frac{1+1}{2} = 1$$

ثانيا المتغير الكمي المتصل

- **تعيين الوسيط حسابياً**

ويمكن أيضاً حساب قيمة الوسيط في حالة التوزيعات التكرارية للمتغيرات المتصلة، وذلك بأن نحدد أولا فئة الوسيط وهى الفئة التي يقع فيها الوسيط. ولإيجاد الوسيط في هذه الحالة نتبع الخطوات الآتية:

(أ) تكوين جدول التكرار المتجمع الصاعد (الهابط).

(ب) تحديد موقع الوسيط وهو في هذه الحالة يساوى $\frac{n}{2}$.

(ج) تحديد الفئة التي تحتوي الوسيط باستخدام موقع الوسيط والتكرار المتجمع الصاعد(الهابط).

(د) إيجاد قيمة الوسيط من العلاقة التالية:

$$Med = L + \left(\frac{\frac{n}{2} - F_1}{F_2 - F_1} \right) \times h$$

حيث L تمثل الحد الأدنى الفعلي لفئة الوسيط، h طول فئة الوسيط، F_1 التكرار المتجمع الصاعد(الهابط) السابق للقيمة $\frac{n}{2}$، F_2 التكرار المتجمع الصاعد اللاحق للقيمة $\frac{n}{2}$.

مثال (٣-١٤)

الجدول الآتي يبين توزيع 100 من الأفراد حسب كمية البروتين (بالجرام) التي يتناولونها يومياً والمطلوب إيجاد الوسيط حسابياً.

كمية البروتين المتناولة	20–	25–	30–	35–	40–	45–	50–	55–
عدد الأفراد	6	14	16	17	14	19	8	6

الحل

المتغير هنا هو كمية البروتين المتناولة وهو متغير كمي متصل ولتعيين الوسيط يجب أولاً حساب التكرار المتجمع الصاعد:

الحدود العليا للفئات	التكرار المتجمع الصاعد F
أقل من 20	0
أقل من25	6
أقل من 30	20
أقل من 35	36
أقل من 40	53
أقل من 45	67
أقل من 50	86
أقل من 55	94
أقل من 60	100

موقع الوسيط:

$$\frac{n}{2}=\frac{100}{2}=50$$

نبحث في التكرار المتجمع الصاعد عن القيمة $\frac{n}{2}=50$ فنجدها غير موجودة ولكنها تقع بين القيمتين 36, 53 لذا فإن الفئة التي تحتوى الوسيط هي 40 – 35 ومنها نجد أن

$$L = 35, \qquad F_1 = 36, F_2 = 53, \qquad h = 5$$

وبالتعويض في القانون فإن

$$Med = L + \left(\frac{\frac{n}{2} - F_1}{F_2 - F_1}\right) \times h$$

$$= 35 + \left(\frac{50 - 36}{53 - 36}\right) \times 5$$

$$= 35 + \frac{14}{17} \times 5 = 39.1176$$

ويمكن بنفس الطريقة تعيين الوسيط لكن باستخدام التكرار المتجمع الهابط.

- **تعيين الوسيط بيانياً**

يمكن تعيين الوسيط بيانياً عن طريق

١- حساب جدول التوزيع التكراري المتجمع الصاعد/الهابط.

٢- رسم منحنى التكرار المتجمع الصاعد/الهابط.

٣- حساب القيمة $\frac{n}{2}$ ثم من المحور الرأسي نحدد موضع تلك القيمة.

٤- نقوم بإسقاط خط أفقي من هذه القيمة على المنحنى ليتقاطع معه في نقطة فنقوم بإسقاط خط رأسي منها ليتقاطع مع المحور الأفقي في نقطة فتكون هي الوسيط.

ويمكن أيضا تعيين الوسيط عن طريق:

١- رسم كل من منحنى التكرار المتجمع الصاعد والهابط(النازل) على نفس المحاور.

٢- من نقطة تقاطع المنحنين نقوم بإسقاط خط رأسي على المحور الأفقي نقطة تلاقى هذا الخط مع الأفقي هي الوسيط.

وبالعودة لمثال (٣-١٤) فيمكن تعيين الوسيط بيانياً كالتالي:

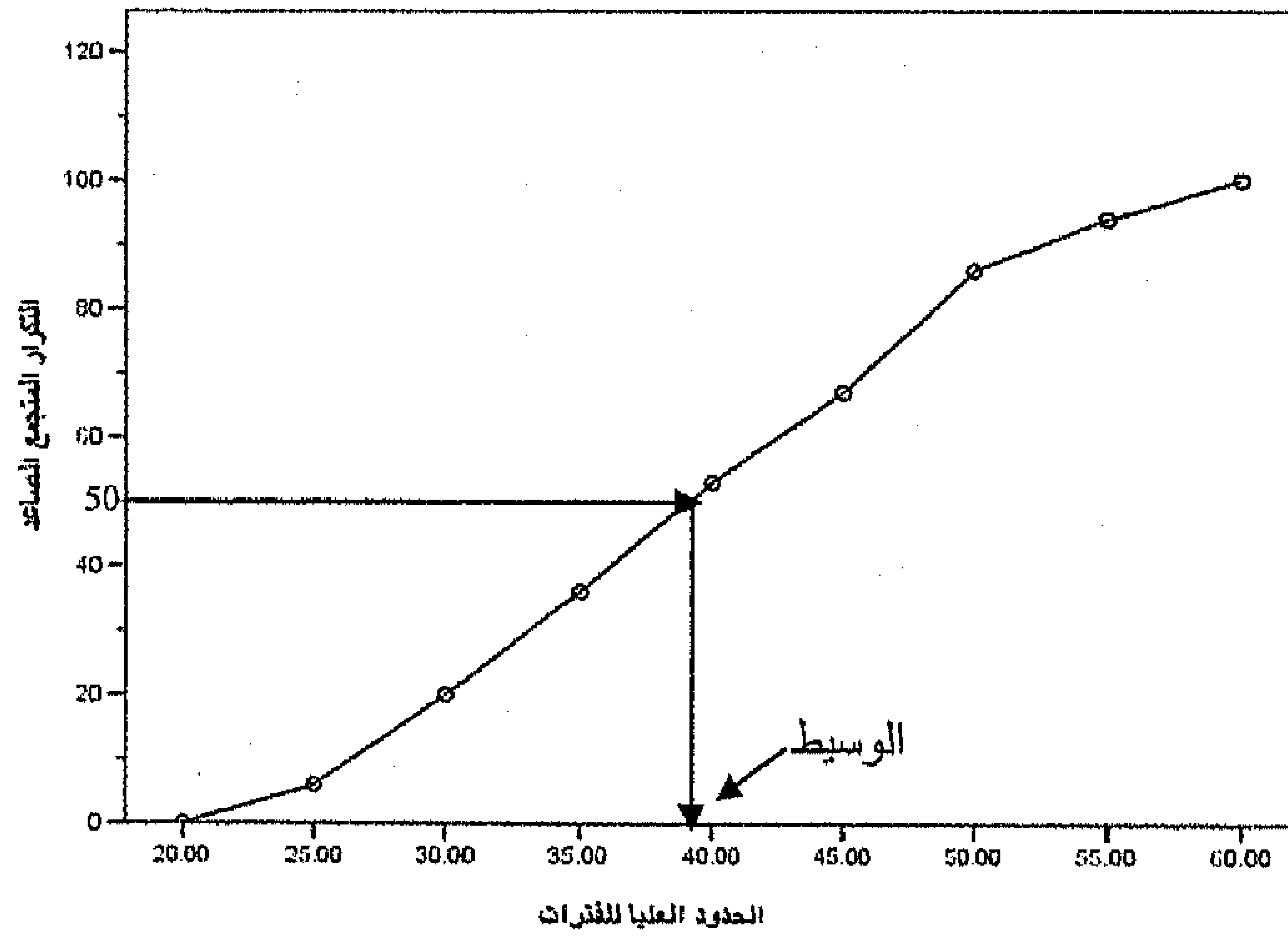
أولا برسم التكرار المتجمع الصاعد

التكرار المتجمع الصاعد F	الحدود العليا للفئات
0	أقل من 20
6	أقل من25
20	أقل من 30
36	أقل من 35
53	أقل من 40
67	أقل من 45
86	أقل من 50
94	أقل من 55
100	أقل من 60

رتبة الوسيط هي

$$\frac{n}{2} = \frac{100}{2} = 50$$

المضلع التكراري المتجمع الصاعد



شكل (٣-١). المضلع التكراري المتجمع الصاعد (مثال ٣-١٤)

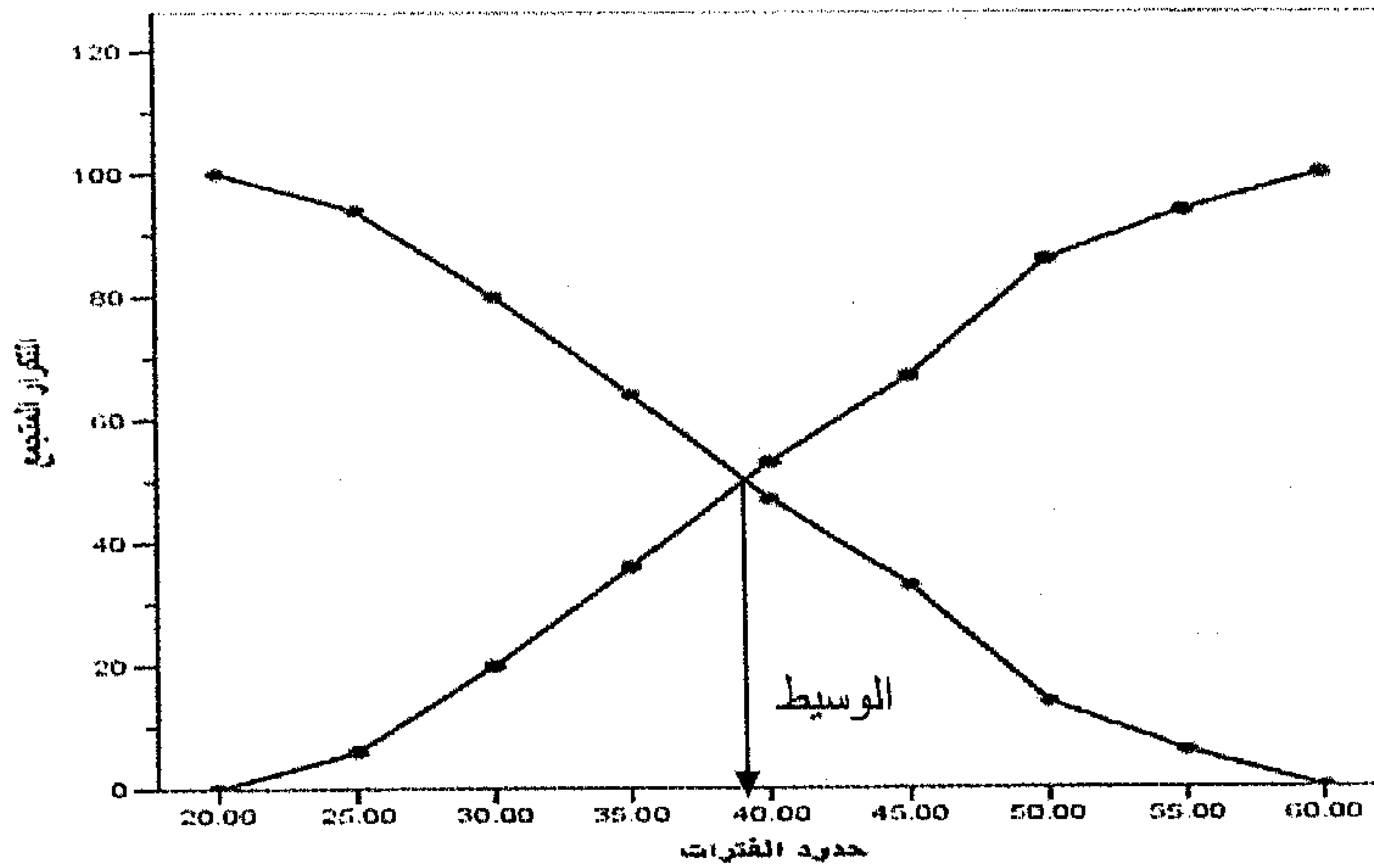
من المضلع التكراري المتجمع الصاعد فإن الوسيط هو

$$Med = 39.1$$

ثانيا برسم التكرار المتجمع الصاعد والهابط معاً:

التكرار المتجمع الهابط	الحدود العليا للفئات
100	20 فأكثر
94	25 فأكثر
80	30 فأكثر
64	35 فأكثر
47	40 فأكثر
33	45 فأكثر
14	50 فأكثر
6	55 فأكثر
0	60 فأكثر

المضلع التكراري المتجمع الصاعد والمضلع التكراري المتجمع الهابط



شكل (٣-٢). المضلع التكراري المتجمع الصاعد والمضلع التكراري المتجمع النازل.

من نقطه تقاطع المنحنين وبعد عملية الإسقاط على الأفقي فإن الوسيط هو

$$Med = 39.1$$

مميزات الوسيط:

يعد الوسيط من مقاييس النزعة المركزية الشائعة، وذلك لوجود بعض الصفات الجيدة له ومن مميزات الوسيط نذكر منها:

١- الوسيط سهل التعريف والحساب.

٢- الوسيط وحيد لمجموعة البيانات الواحدة.

٣- الوسيط أقل تأثراً من المتوسط بالقيم المتطرفة والشاذة.

عيوب الوسيط:

بالرغم من أن الوسيط يعتبر من مقاييس النزعة المركزية الجيدة إلا أن فيه بعض العيوب نذكر منها ما يلي:

١- لا يأخذ في الاعتبار جميع البيانات لأنه يعتمد فقط على القيمة التي في المنتصف.

٢- لا يمكن حساب الوسيط للبيانات الوصفية(النوعية) بشكل عام.

ملاحظة(٣-٣):

وحدة قياس الوسيط هي نفس وحدة البيانات الأصلية. فإذا كانت وحدة البيانات بالجرام فإن وحدة الوسيط هي الجرام.

٣-٨ الربيعات والعشيرات والمئينات (Quartiles, Decimals and Percentages)

وكما سبق يمكن بنفس الطريقة التي تم بها حساب الوسيط حسابياً وأيضاً بيانياً يمكن تعيين كل من الربيعات والعشيرات وأيضا المئينات؛ فالربيعات هي مقاييس تقسم البيانات لأربعة أجزاء متساوية ولذلك يوجد ثلاثة ربيعات وهى الربيع الأول Q_1 والذى يوجد قبله 25% من البيانات وبعده 75% من البيانات والربيع الثاني والذي يوجد قبله 50% من البيانات وبعده أيضا 50% من البيانات، والربيع الثالث والذي يوجد قبله 75% من البيانات وبعده 25% من البيانات ومن الملاحظ أن الربيع الثاني هو الوسيط للبيانات.

ويمكن حساب الربيعات في حالة البيانات الكمية المتصلة من القانون التالي:

$$Q_i = L + \left(\frac{R_i - F_1}{F_2 - F_1}\right) \times h\,, \qquad i = 1, 2, 3$$

حيث $R_i = \frac{in}{4}$ هي رتبة الربيع رقم i والتي باستخدامها يمكن تحديد موقع الربيع المراد تعيينه،

F_1: التكرار المتجمع السابق لرتبه الربيع،

F_2: التكرار المتجمع اللاحق لرتبة الربيع،

h: طول الفئة التي تحتوى الربيع،

L: الحد الأدنى الفعلي للفئة التي تحتوى الربيع.

مثال (٣-١٥)

فيما يلي التوزيع التكراري لعدد الحوادث اليومية خلال 100 يوم متتالية على احد الطرق:

X عدد الحوادث اليومية	0	1	2	3	4	5
f عدد الأيام	48	32	17	2	1	0

أوجد الربيع الأول والثاني والثالث لعدد الحوادث اليومية.

الحل

نلاحظ أن X متغير كمي متقطع (منفصل) وأن إجمالي عدد الأيام n = 100 والتكرار المتجمع الصاعد هو

عدد الحوادث اليومية x_i	عدد الأيام f_i	التكرار المتجمع الصاعد
0	48	48
1	**32**	**80**
2	17	97
3	2	99
4	1	100
5	0	100

الربيع الأول:

رتبه الربيع الأول هي

$$R_1 = 1 \times \frac{n}{4} = \frac{100}{4} = 25$$

من التكرار المتجمع الصاعد نجد أن القيمة 25 غير موجودة، لذا سوف نبحث عن القيمة التالية الموجودة في الجدول وهى 48. فإن القيمة التي يقابلها تكرار متجمع صاعد يساوى 48 هي القيمة 0 فإن

$$Q_1 = 0$$

الربيع الثاني:

رتبه الربيع الثاني هو

$$R_2 = \frac{2 \times n}{4} = \frac{2 \times 100}{4} = 50$$

من التكرار المتجمع الصاعد نجد أن القيمة 50 غير موجودة، لذا سوف نبحث عن القيمة التالية الموجودة في الجدول وهى 80. فإن القيمة التي يقابلها تكرار متجمع صاعد يساوى 80 هي القيمة 1 فإن

$$Q_2 = 1$$

ونلاحظ أن قيمة الربيعي الثاني تساوى قيمة الوسيط الذي تم تعيينه في مثال(٣-١٣) سابقاً.

الربيع الثالث:

رتبه الربيع الثالث هي

$$R_3 = \frac{3 \times n}{4} = \frac{3 \times 100}{4} = 75$$

من التكرار المتجمع الصاعد نجد أن القيمة 75 غير موجودة، لذا سوف نبحث عن القيمة التالية الموجودة في الجدول وهى 80. فإن القيمة التي يقابلها تكرار متجمع صاعد يساوى 80 هي القيمة 1 فإن

$$Q_3 = 1$$

مثال (٣-١٦)

فيما يلي التوزيع التكراري لإنتاج مجموعة من قطع الأرض الزراعية المتساوية من أحد المحاصيل(بآلاف الكيلوجرام):

فئات الإنتاج	26-	31-	36-	41-	46-	51-	56-	61-
عدد القطع	4	5	23	58	61	30	4	3

احسب كلاً من الربيع الأول والثالث حسابياً وبيانياً.

الحل

بتعيين جدول التكرار المتجمع الصاعد

التكرار المتجمع الصاعد	الحدود العليا للفئات
0	أقل من 26
4	أقل من31
9	أقل من 36
32	أقل من 41
90	أقل من 46
151	أقل من 51
181	أقل من 56
185	أقل من 61
188	أقل من 66

أولا حسابياً

- **الربيع الأول**

موقع(رتبة) الربيع الأول

$$R_1 = \frac{n}{4} = \frac{188}{4} = 47$$

نبحث في التكرار المتجمع الصاعد عن القيمة $R_1 = 47$ فنجدها غير موجودة، ولكنها تقع بين القيمتين 32, 90 الموجودتين في الجدول فإن الفترة(الفئة) التي تحتوى الربيع الأول هي الفترة 46–41 ومنها فإن

$$L = 41, F_1 = 32, F_2 = 90, h = 5$$

وبالتعويض في القانون

$$Q_1 = L + \left(\frac{R_1 - F_1}{F_2 - F_1}\right) \times h$$

$$= 41 + \left(\frac{47 - 32}{90 - 32}\right) \times 5 = 42.293$$

- **الربيع الثالث**

موقع الربيع الثالث هو

$$R_3 = \frac{3 \times n}{4} = \frac{3 \times 188}{4} = 141$$

نبحث في التكرار المتجمع الصاعد عن القيمة $R_3 = 141$ فنجدها غير موجودة ولكنها تقع بين القيمتين 90, 151 الموجودتين في الجدول. فإن الفترة(الفئة) التي تحتوى الربيع الأول هي الفترة 51-46 ومنها فإن

$$L = 46, \quad h = 5, \quad F_1 = 90, \quad F_2 = 151$$

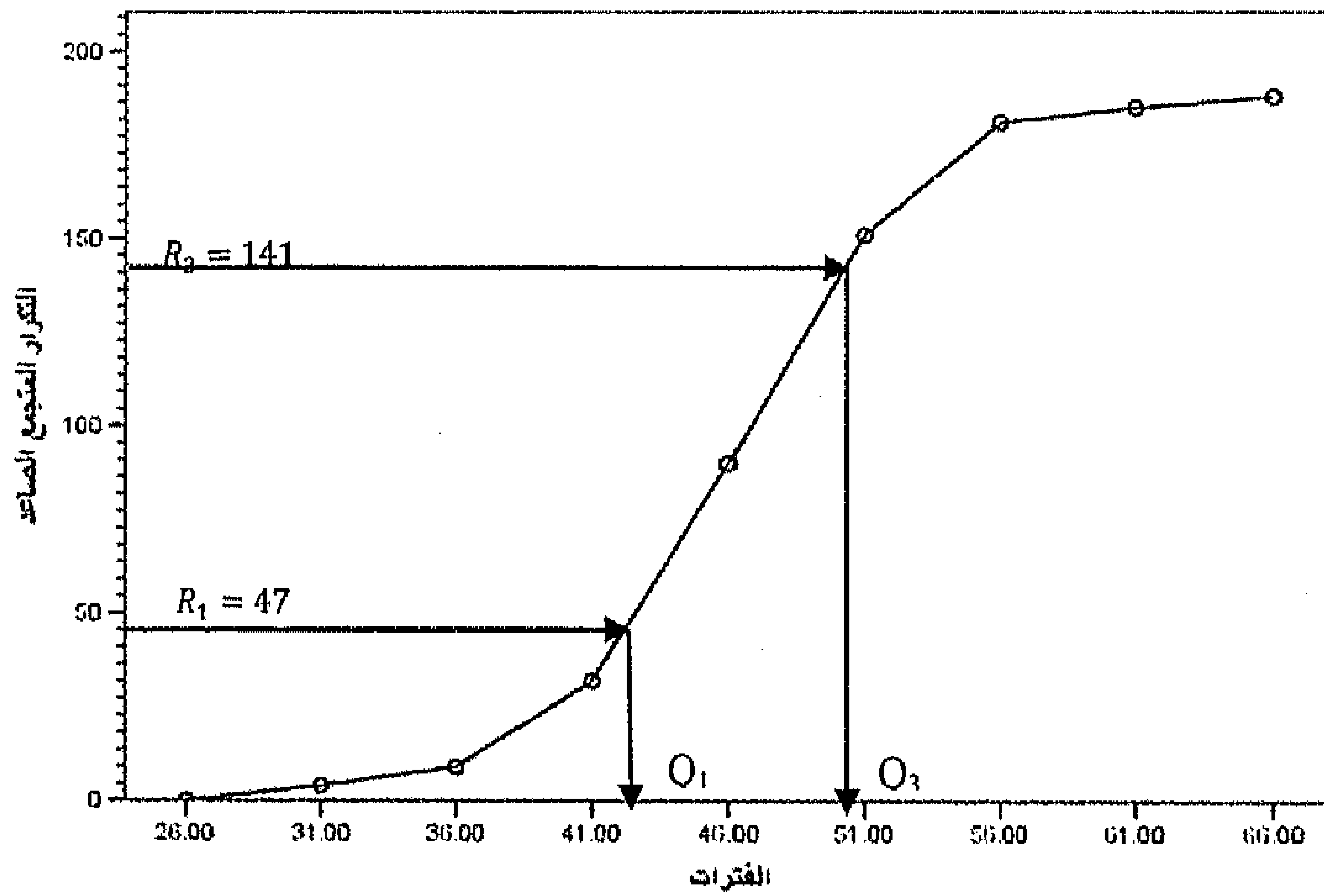
وبالتعويض في القانون

$$Q_3 = L + \left(\frac{R_3 - F_1}{F_2 - F_1}\right) \times h$$

$$= 46 + \left(\frac{141 - 90}{151 - 90}\right) \times 5 = 50.180$$

ثانيا بيانياً

برسم المضلع التكراري المتجمع الصاعد



شكل (٣-٣). المضلع التكراري المتجمع الصاعد.

- **الربيع الأول**

موقع(رتبة) الربيع الأول

$$R_1 = \frac{n}{4} = \frac{188}{4} = 47$$

من المحور الرأسي (محور y) الممثل للتكرار المتجمع الصاعد نبحث عن القيمة $R_1 = 47$ نقوم بإسقاط خط أفقي من تلك النقطة فيقابل منحنى التكرار المتجمع الصاعد عند نقطة منها نقوم مره ثانية بإسقاط خط رأسي ليقابل المحو الأفقي (الفترات) عند نقطة فتكون هي قيمة الربيع الأول ونجد أن:

$$Q_1 = 42.3$$

- **الربيع الثالث**

موقع الربيع الثالث هو

$$R_3 = \frac{3 \times n}{4} = \frac{3 \times 188}{4} = 141$$

من المحور الرأسي (محور y) الممثل للتكرار المتجمع الصاعد نبحث عن القيمة $R_3 = 141$ ونقوم بإسقاط خط أفقي من تلك النقطة فيقابل منحنى التكرار المتجمع الصاعد عند نقطة منها نقوم مره ثانية بإسقاط خط رأسي ليقابل المحور الأفقي (الفترات) عند نقطة فتكون هي قيمة الربيع الثالث ونجد أن

$$Q_3 = 50.2$$

وكما سبق يمكن بنفس الطريقة تعيين العشيرات، فالعشيرات هي مقاييس تقسم البيانات لعشرة أجزاء متساوية ولذلك يوجد تسعة عشيرات؛ وهى العشير الأول D_1 والذى يوجد قبله $\frac{1}{10}$ من البيانات وبعده $\frac{9}{10}$ من البيانات والعشير الثاني والذي يوجد قبله $\frac{2}{10}$ من البيانات وبعده ايضا $\frac{8}{10}$ من البيانات،... والعشير التاسع والذي يوجد قبله $\frac{9}{10}$ من البيانات وبعده $\frac{1}{10}$ من البيانات. ومن الملاحظ أن العشير الخامس هو الوسيط للبيانات.

ويمكن حساب العشيرات في حالة البيانات الكمية المتصلة من القانون التالي:

$$D_i = L + \left(\frac{R_i^* - F_1}{F_2 - F_1}\right) \times h\,, \qquad i = 1,2,3,\ldots,9$$

حيث $R_i^* = \frac{in}{10}$ هي رتبة العشير رقم i والتي باستخدامها يمكن تحديد موقع العشير المراد تعيينه،

F_1: التكرار المتجمع السابق لرتبه العشير،

F_2: التكرار المتجمع اللاحق لرتبة العشير،

h: طول الفئة التي تحتوى العشير.

L: الحد الأدنى الفعلي للفئة التي تحتوى العشير.

وأيضا يمكن تعريف المئينات على أنها هي مقاييس تقسم البيانات لمائة جزء متساوي، ولذلك يوجد تسع وتسعين مئين وهى المئين الأول P_1 والذى يوجد قبله $\frac{1}{100}$ من البيانات وبعده $\frac{99}{100}$ من البيانات والمئين الثاني والذي يوجد قبله $\frac{2}{100}$ من البيانات وبعده ايضا $\frac{98}{100}$ من البيانات،... والمئين التاسع والتسعين والذي يوجد قبله $\frac{99}{100}$ من البيانات وبعده $\frac{1}{100}$ من البيانات، ومن الملاحظ أن المئين الخمسين هو الوسيط للبيانات.

ويمكن حساب المئينات في حالة البيانات الكمية المتصلة من القانون التالي:

$$P_i = L + \left(\frac{R_i^{**} - F_1}{F_2 - F_1}\right) \times h , \qquad i = 1, 2, 3, \dots, 99$$

حيث $R_i^{**} = \frac{in}{100}$ هي رتبة المئين رقم i والتي باستخدامها يمكن تحديد موقع المئين المراد تعيينه،

F_1: التكرار المتجمع السابق لرتبه المئين،

F_2: التكرار المتجمع اللاحق لرتبة المئين،

h: طول الفئة التي تحتوى المئين،

L: الحد الأدنى الفعلي للفئة التي تحتوى المئين.

مثال (٣-١٧)

فيما يلي التوزيع التكراري لإنتاج (بآلاف الكيلوجرام) مجموعات من قطع الأرض الزراعية المتساوية من أحد المحاصيل:

فئات الإنتاج	26–	31–	36–	41–	46–	51–	56–	61–
عدد القطع	4	5	23	58	61	30	4	3

احسب

(أ) العشير السابع

(ب) المئين الثالث والسبعين

الحل:

بتعيين جدول التكرار المتجمع الصاعد

التكرار المتجمع الصاعد	الحدود العليا للفئات
0	أقل من 26
4	أقل من31
9	أقل من 36
32	أقل من 41
90	أقل من 46
151	أقل من 51
181	أقل من 56
185	أقل من 61
188	أقل من 66

(أ) العشير السابع

موقع(رتبة) العشير السابع

$$R_7^* = \frac{7 \times n}{10} = \frac{7 \times 188}{10} = 131.6$$

نبحث في التكرار المتجمع الصاعد عن القيمة $R_7^* = 131.6$ فنجدها غير موجودة ولكنها تقع بين القيمتين 90 ,151 الموجودتين في الجدول فإن الفترة(الفئة) التي تحتوى العشير السابع هي الفترة 46–51، ومنها فإن

$$L = 46, F_1 = 90, F_2 = 151, h = 5$$

وبالتعويض في القانون

$$D_7 = L + \left(\frac{R_7^* - F_1}{F_2 - F_1}\right) \times h$$

$$= 46 + \left(\frac{131.6 - 90}{151 - 90}\right) \times 5 = 49.41$$

(ب) المئين الثالث والسبعين

موقع المئين الثالث والسبعين هو

$$R_{73}^{**} = \frac{73 \times n}{100} = \frac{73 \times 188}{100} = 137.24$$

نبحث في التكرار المتجمع الصاعد عن القيمة $R_{73}^{**} = 137.24$ فنجدها غير موجودة ولكنها تقع بين القيمتين 90 ,151 الموجودتين في الجدول فإن الفترة(الفئة) التي تحتوى المئين الثالث والسبعين هي الفترة 46–51، ومنها فإن

$$L = 46, F_1 = 90, F_2 = 151, h = 5$$

وبالتعويض في القانون

$$P_{73} = L + \left(\frac{R_{73}^{**} - F_1}{F_2 - F_1}\right) \times h$$

$$= 46 + \left(\frac{137.24 - 90}{151 - 90}\right) \times 5 = 49.872$$

ويمكن تعيين العشيرات والمئينات بيانياً كما مع الوسيط والربيعيات ومما سبق نجد أن:

$$Q_2 = D_5 = P_{50} = Med$$

٣-٩ المنوال (Mode)

المنوال من مقاييس النزعة المركزية التي لها أهمية في بعض الدراسات. ويعرف المنوال على أنه القيمة الأكثر تكراراً (ظهوراً) بين المشاهدات. ويمكن تعيين المنوال بسهوله ويسر للبيانات المفردة عن طريق تحديد القيمة الأكثر تكراراً فتكون هي المنوال، فالمنوال للمشاهدات 3 ,9 ,7 , 6 , 5 ,6 ,4 ,6 ,5 هو 6 لأنها الأكثر تكراراً بينها.
لكن يعيب المنوال أنه قد لا يوجد منوال في حالة البيانات التي تتساوى فيها تكرارات المشاهدات أو أنه يوجد أكثر من منوال، إذا كان أكبر تكرار يتحقق لقيمتين أو أكثر من المشاهدات معاً.

البيانات المبوبة (Grouped Data):

تعيين المنوال حسابياً:

بالنسبة للبيانات الكمية المنفصلة فإننا سوف نبحث في التكرار عن أكبر قيمة وسيكون المنوال هي القيمة التي تقابل أكبر تكرار لكن بالنسبة للبيانات الكمية المتصلة فإننا سوف نتبع الخطوات التالية لتعيين المنوال:

(أ) نبحث في التكرار عن أكبر تكرار f

(ب) نحدد فئة المنوال،

(ج) نحدد التكرار السابق لفئة المنوال f_1 والتكرار اللاحق لفئة المنوال f_2

(د) نعين المنوال من القانون

$$Mod = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times h$$

حيث

L هو الحد الأدنى الفعلي لفئة المنوال، h طول فئة المنوال،

Δ_1 الفرق بين أكبر تكرار والتكرار السابق له أي أن

$$\Delta_1 = f - f_1$$

Δ_2 الفرق بين أكبر تكرار والتكرار اللاحق له أي أن

$$\Delta_2 = f - f_2$$

ويمكن اعتبار مركز الفئة المنوالية هو منوال البيانات.

مثال (٣-١٨)

في عينة من 80 زجاجة سعة كل منها لتر واحد تم قياس ما تحتويه من سائل بالملي لتر وتم وضع البيانات في الجدول الآتي:

كمية السائل	910−	930−	950−	970−	990−	1010−	1030−	1050−
عدد الزجاجات	1	5	9	20	25	15	4	1

أوجد المنوال لما تحتويه الزجاجة من سائل حسابياً.

الحل

بالبحث في التكرار نجد أن أكبر تكرار هو $f = 25$ فإن الفئة المنوالية أي الفئة التي تحتوى المنوال هي1010−990 ومنها نجد أن

$$L = 990, \qquad h = 20$$

التكرار السابق لأكبر تكرار هو $f_1 = 20$ والتكرار اللاحق لأكبر تكرار هو $f_2 = 15$ فان

$$\Delta_1 = f - f_1 = 25 - 20 = 5$$

$$\Delta_2 = f - f_2 = 25 - 15 = 10$$

سوف يحسب المنوال من العلاقة

$$Mod = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times h$$

$$= 990 + \left(\frac{5}{5 + 10}\right) \times 20 = 996.667$$

تعيين المنوال بيانياً:

يمكن تعيين المنوال بيانياً للبيانات الكمية المتصلة، وذلك برسم المدرج التكراري للبيانات، ونقوم باستخدام أعلى عمود برسم خط مستقيم يوصل بين نهاية العمود مع نهاية العمود السابق له وأيضاً خط مستقيم يوصل بين بداية ذلك العمود مع بداية العمود اللاحق له، من نقطة تقاطع هذين الخطين نقوم بإسقاط خط رأسي يتقابل مع الخط الأفقي (محور X) في نقطة فتكون هي المنوال.

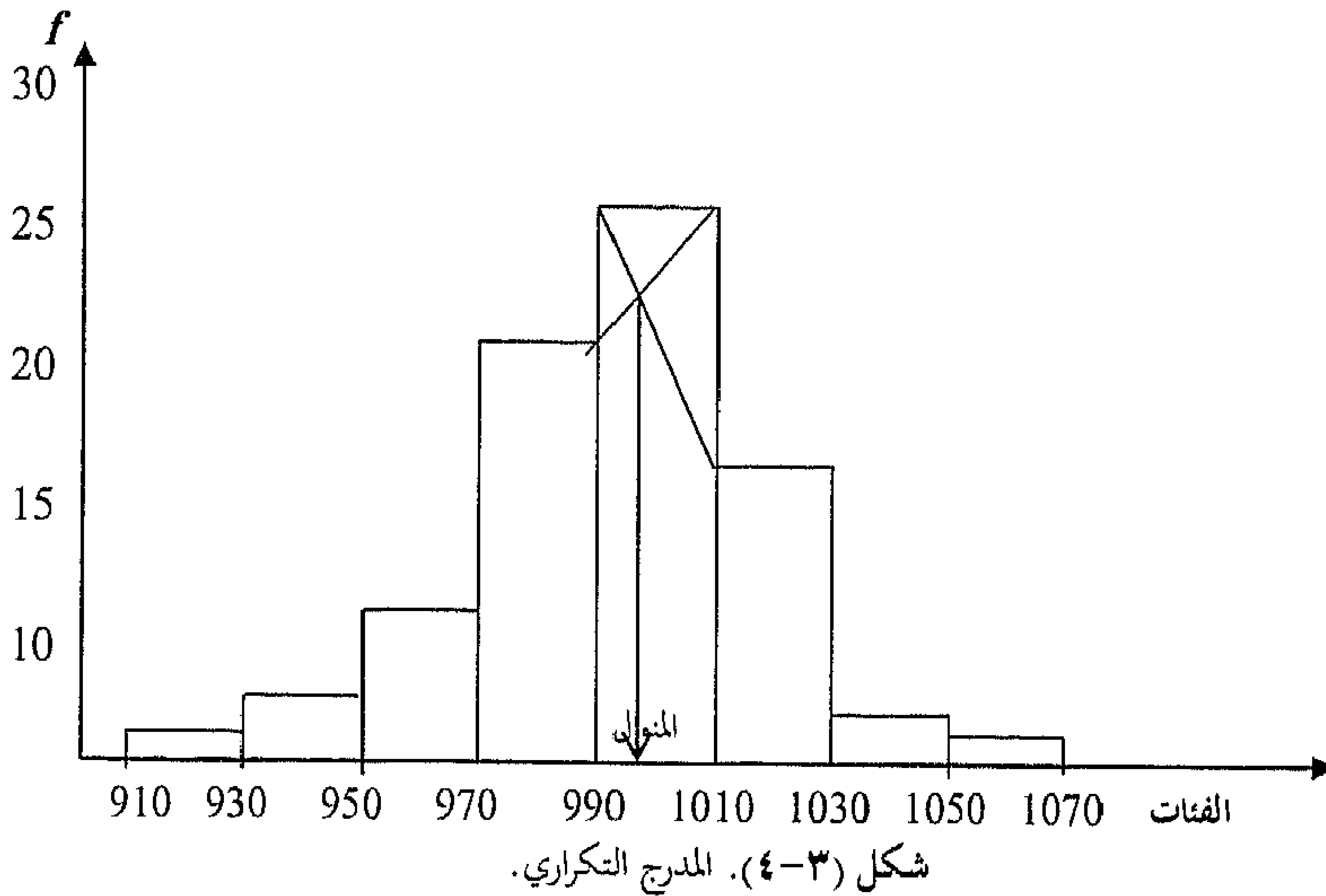
مثال (٣-١٩)

في عينة من 80 زجاجة سعة كل منها لتر واحد تم قياس ما تحتويه من سائل بالملي لتر وتم وضع البيانات في الجدول الآتي:

كمية السائل	910–	930–	950–	970–	990–	1010–	1030–	1050–
عدد الزجاجات	1	5	9	20	25	15	4	1

أوجد المنوال لما تحتويه الزجاجة من سائل بيانياً.

الحل: برسم المدرج التكراري كما بالشكل (٣-٤).



شكل (٣-٤). المدرج التكراري.

فمن الشكل البياني (٣-٤) نجد أن الفئة 1010 - 990 يقابلها أعلى عمود لذا سوف نرسم خط يصل بين نهايته ونهاية العمود المرسوم على الفئة السابقة 990 – 970 ونرسم خطاً مستقيماً يصل بين بداية أعلى عمود وبداية العمود المرسوم على الفئة التالية 1030 – 1010 من نقطة تقاطع الخطين نسقط عموداً ليتقابل مع الخط الأفقي (الفئات) وتكون هي المنوال ونلاحظ أنها أقرب لبداية الفترة ومنها فإن قيمة المنوال هي تقريباً 997

وإذا كانت أطوال الفترات للتوزيع التكراري غير متساوية فإنه يجب إيجاد التكرارات المعدلة وهي عبارة عن التكرارات الحقيقية مقسومة على أطوال الفئات المناظرة لها. وتستخدم التكرارات المعدلة في إيجاد المنوال سواء بيانياً أو حسابياً.

مثال (٣-٢٠)

الجدول الآتي يبين توزيع 98 نباتا حسب أطوالها بالسنتيمتر

فئات الطول	10–	15–	18–	20–	22–	25–30
عدد النباتات	15	24	22	20	12	5

المطلوب إيجاد المنوال لطول النبات.

الحل

لاحظ أن فئات أطوال النباتات غير متساوية. وعلى ذلك يجب حساب التكرار المعدل قبل إيجاد المنوال

فئات الطول	التكرار	طول الفئات	التكرار المعدل
10–	15	5	3
15–	24	3	8
18–	22	2	11
20–	20	2	10
22–	12	3	4
25–30	5	5	1

أكبر تكرار معدل هو 11 فإن فئة المنوال هي 20–18 وحدها الأدنى $L = 18$ وطولها $h = 2$ والفروق

$$\Delta_1 = f - f_1 = 11 - 8 = 3$$

$$\Delta_2 = f - f_2 = 11 - 10 = 1$$

وبالتعويض في القانون فإن المنوال

$$Mod = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times h$$

$$= 18 + \left(\frac{3}{3+1}\right) \times 2 = 19.5$$

مميزات المنوال:

يعد المنوال من مقاييس النزعة المركزية الشائعة الاستخدام ويتمتع ببعض المميزات منها:

١- المنوال سهل التعريف والحساب.

٢- المنوال أقل تأثراً من المتوسط بالقيم الشاذة والمتطرفة.

٣- يمكن حساب المنوال للبيانات النوعية(الوصفية).

عيوب المنوال:

يوجد بعض العيوب للمنوال منها:

١- لا يأخذ المنوال في الاعتبار جميع البيانات حيث يعتمد على القيمة الأكثر تكراراً.

٢- قد لا يوجد منوال لمجموعة من البيانات أو قد يكون هناك أكثر من منوال.

ملاحظة (٣-٤):

وحدة المنوال هي نفس وحدة البيانات الأصلية. فإذا كانت البيانات هي بالكيلوجرام فإن وحدة المنوال هي الكيلوجرام.

٣-١٠ العلاقة بين الوسط الحسابي والوسيط والمنوال

سبق أن ذكرنا في الفصل الثاني أن التوزيع التكراري يمكن أن يكون متماثلاً أو ملتوياً جهة اليمين أو ملتوياً جهة اليسار(الشمال) شكل (٢-٩). فإذا كان

١. التوزيع متماثلاً (symmetric) فإن:

$$\bar{x} = Median = Mode$$

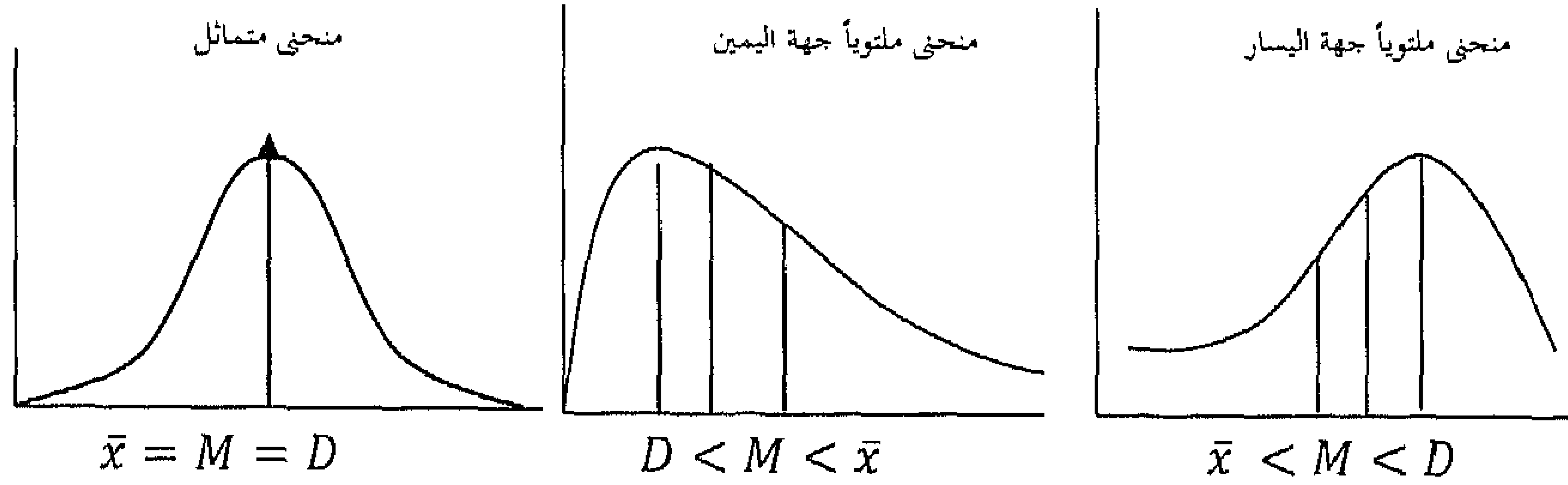
٢. التوزيع ملتوياً جهة اليمين فإن:

$$Mode < Median < \bar{x}$$

٣. التوزيع ملتوياً جهة اليسار فإن:

$$\bar{x} < Median < Mode$$

وفى الحالتين الأخيرتين نجد دائماً أن الوسيط يقع بين المنوال والوسط الحسابي. كما بالشكل التالي:



شكل(٣-٥). العلاقة بين الوسط والمنوال والوسيط.

حيث D تمثل المنوال، M الوسيط، $\bar{x}$ تمثل الوسط الحسابي وقد لوحظ أنه كلما أقترب التوزيع من التماثل فإن

$$\bar{x} - Mod = 3(\bar{x} - Med)$$

٣-١١ تطبيقات باستخدام برنامج SPSS

سوف نقوم بإجراء بعض التطبيقات للتعرف على كيفية حساب مقاييس النزعة المركزية، وذلك بحل بعض التمارين التي تم التطرق لها سابقاً ولكن باستخدام برنامج العرض والتحليل الإحصائي SPSS وسوف نقوم بتعيين جميع المقاييس معاً وليس كل مقياس على حده.

البيانات المفردة للمتغير الكمي المنفصل

تطبيق (٣-١):

من مثال(٣-١) فيما يلي أوزان 12 طالباً من طلاب المستوى الأول بإحدى كليات جامعة الخرج (بالكيلو جرام).

64 78 91 58 81 69 103 65 51 77 76 63

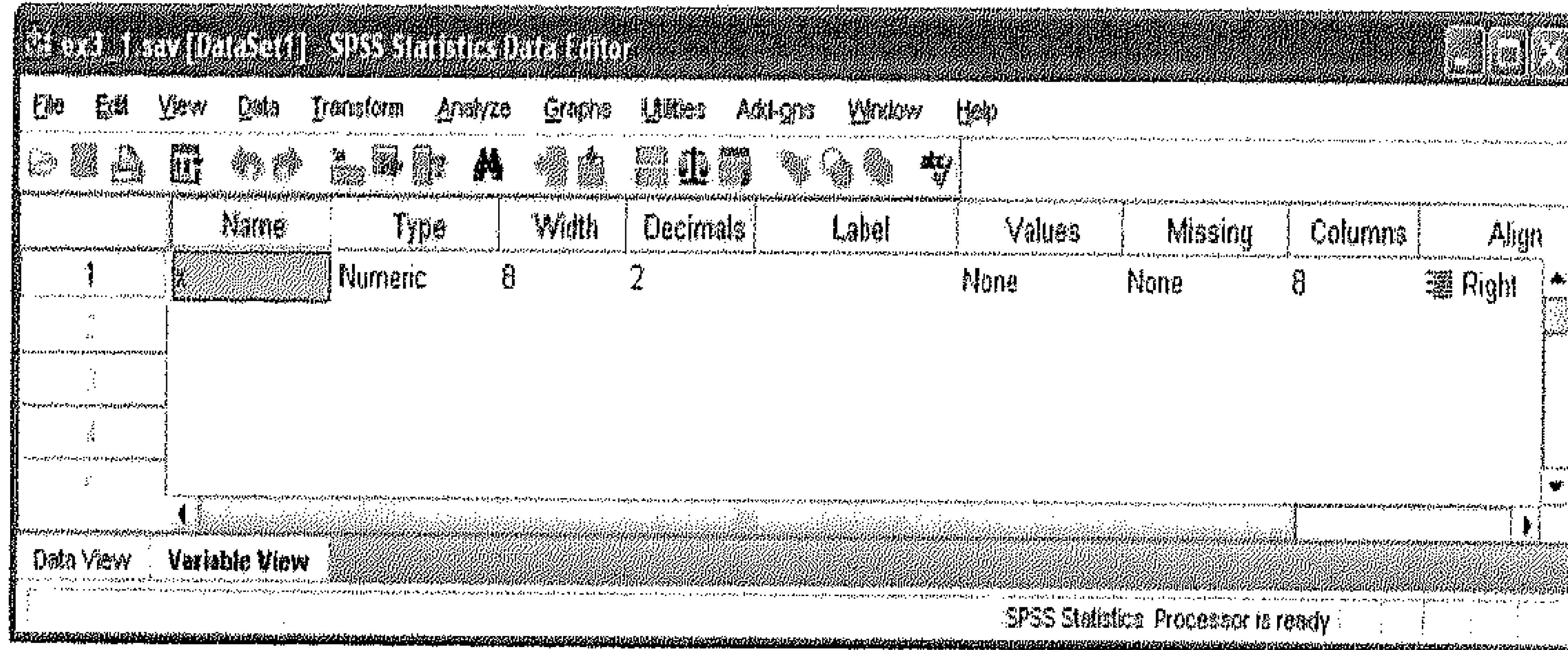
أوجد مقاييس النزعة المركزية لأوزان هؤلاء الطلاب.

الحل

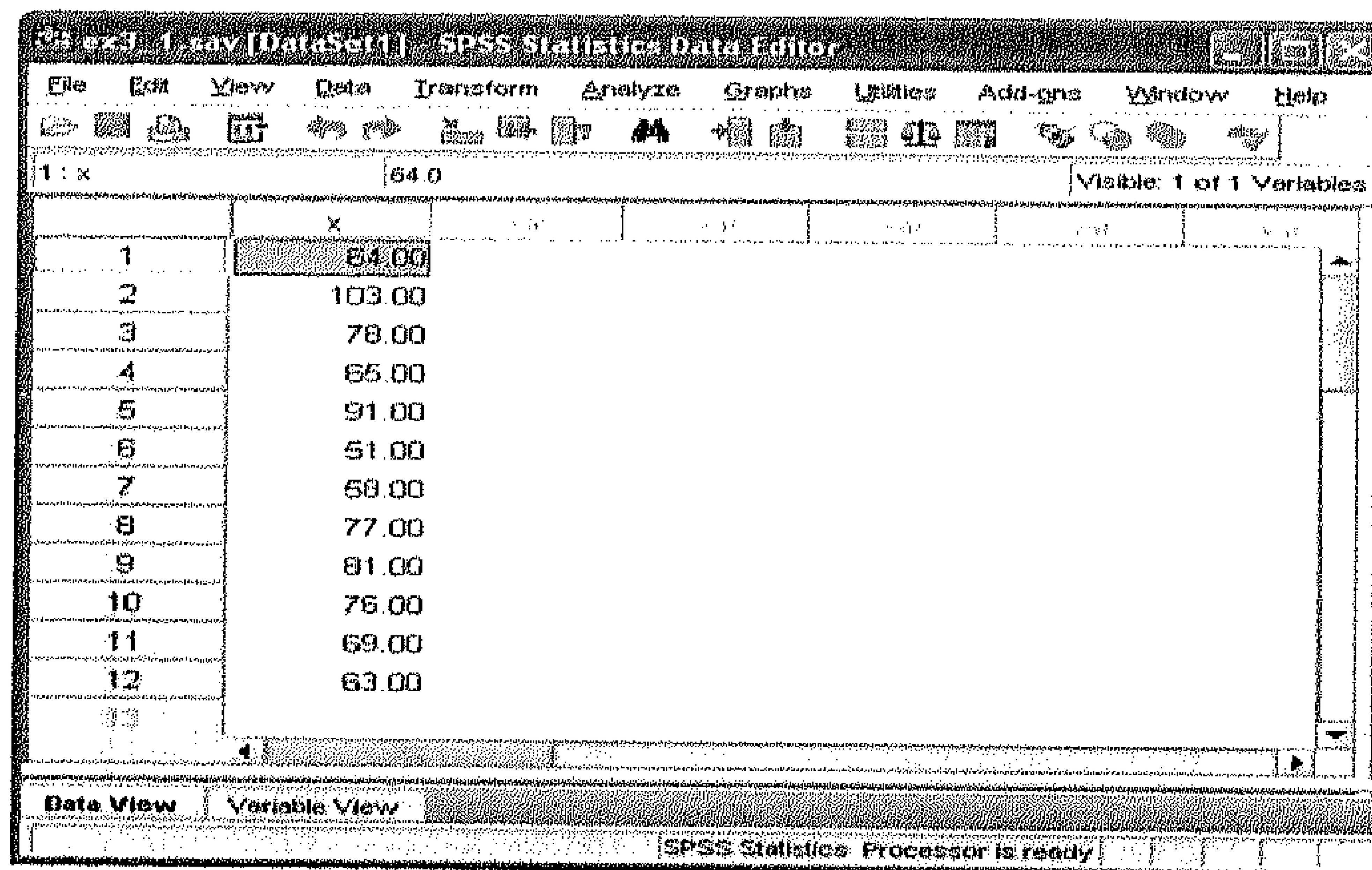
أولا نقوم بإدخال البيانات للبرنامج بإتباع الخطوات التالية:

١- بعد فتح ملف جديد نضغط على Variable View الموجودة أسفل يمين نافذة البرنامج

٢- من نافذة Variable View نقوم بتعريف متغير X وهو متغير كمي

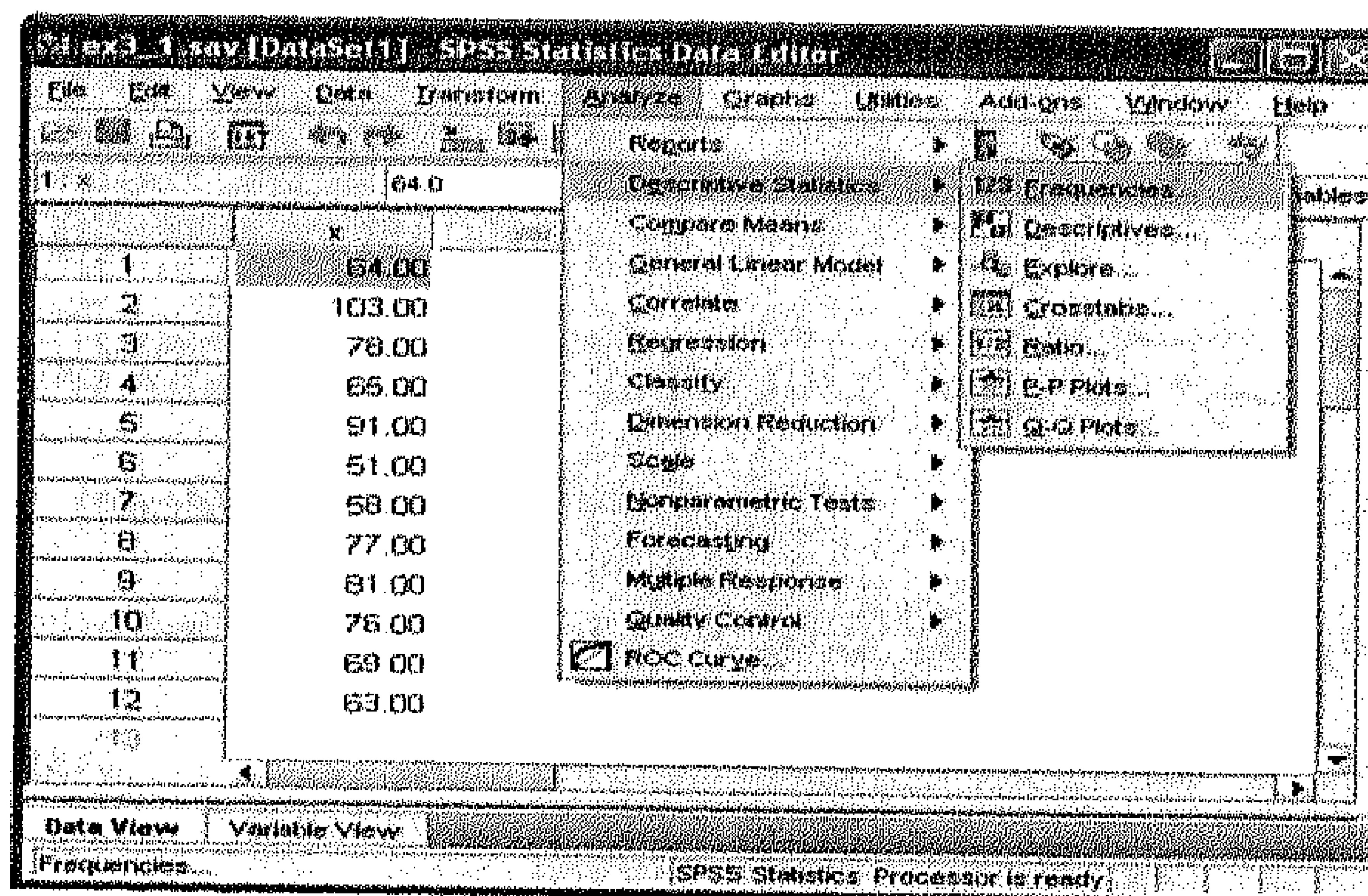


٣- ننتقل لشاشة Data View ونقوم بإدخال البيانات للمتغير X

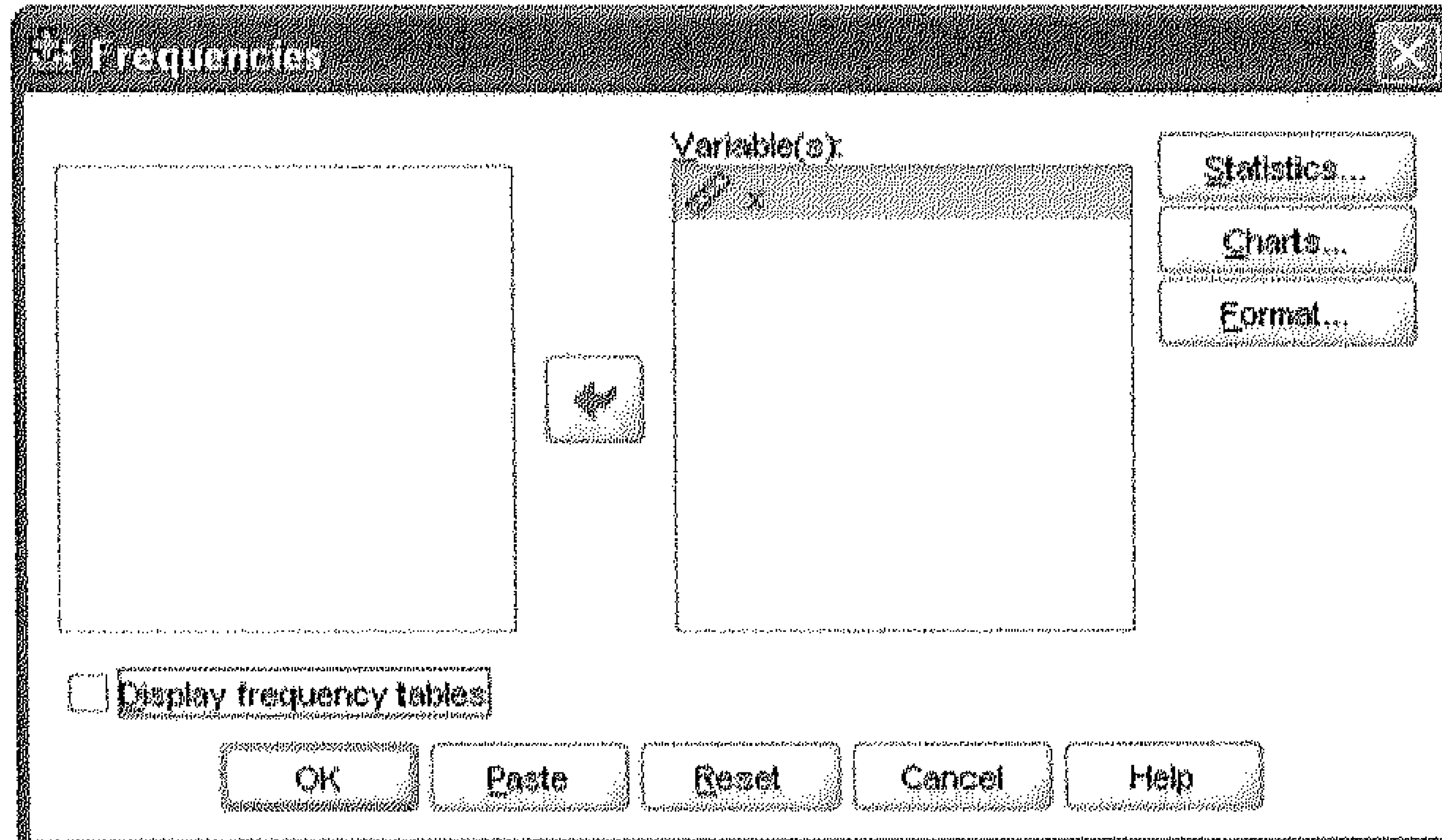


ثانياً تعيين المقاييس المطلوبة بإتباع الخطوات التالية:

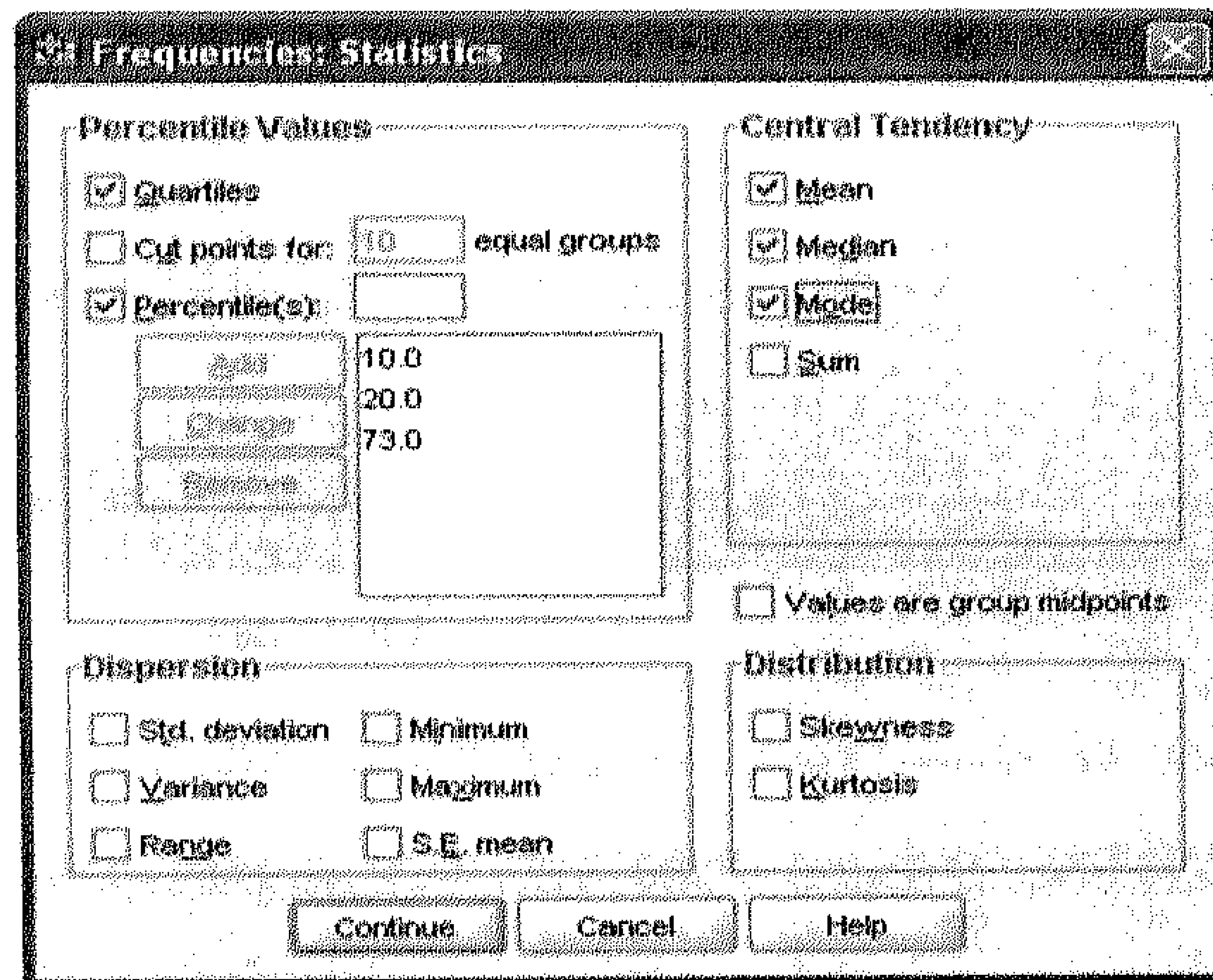
١- من قائمة Analyze نختار Descriptive Statistics ومن القائمة المنسدلة نختار Frequencies



٢- تظهر شاشة جديدة بعنوان Frequencies ننقل المتغير X لقائمة Variables لاحظ أن الاختيار Display frequency tables غير محدد لأننا لا نرغب في عرض الجدول التكراري.



٣- نضغط على الاختيار Statistics تظهر شاشة تحتوى على عديد من المقاييس الإحصائية



سوف نهتم هنا بكل من

- قائمة مقاييس النزعة المركزية (Central Tendency) التي تحتوى على الوسط (Mean) والوسيط (Median) والمنوال (Mode) والمجموع (Sum)
- قائمة Percentile values التي تستخدم لتعيين الربيعات (Quartiles) والعشيرات والمئينات باستخدام الاختيار Percentiles

٤- من قائمة Central Tendency سوف نحدد كلا من Mean لتعيين المتوسط، Median لتعيين الوسيط، Mode لتعيين المنوال.

٥- من قائمة Percentile Values نختار Quartiles لتعيين الربيعات ونختار Percentile(s) ونكتب 10 وتعنى القيمة التي قبلها 10% من البيانات وهى تمثل العشير الأول وأيضا المئين العاشر، ثم نضغط على Add ثم نكتب 20 وتمثل العشير الثاني وأيضا المئين العشرين، نكتب 73 وتمثل المئين الثالث والسبعين.

٦- نضغط Continue فنعود للشاشة السابقة نضغط Ok فتظهر النتائج التالية:

Statistics

x

N	Valid	12
	Missing	0
Mean		73.0000
Median		72.5000
Mode		51.00[a]
Percentiles	10	53.1000
	20	61.0000
	25	63.2500
	50	72.5000
	73	79.4700
	75	80.2500

a. Multiple modes exist. The smallest value is shown

من الجدول السابق نجد أن

- عدد القيم المتاحة للمتغير هي $n = 12$ ، عدد القيم الشاذة (Missing) هي 0
- الوسط الحسابي للبيانات هو $\bar{x} = 73$ ، الوسيط للبيانات هو $Med = 72.5$
- المنوال للبيانات هو $Mod = 51.0$، المئين العاشر(العشير الأول) هو $P_{10} = D_1 = 53.1$
- المئين العشرين (العشير الثاني) قيمته $P_{20} = D_2 = 61.0$، الربيع الأول قيمته هي $Q_1 = 63.25$
- الربيع الثاني قيمته هي $Q_2 = 72.5$ ونجد أنه نفس قيمة الوسيط حيث إنه من المعلوم أن الربيع الثاني يساوى الوسيط، المئين الثالث والسبعين قيمته هي $P_{73} = 79.47$ ، الربيع الثالث قيمته $Q_3 = 80.25$

البيانات المبوبة للمتغير الكمي المنفصل:

تطبيق (٣-٢):

في مثال (٣-٣) أوجد مقاييس النزعة المركزية لعدد حبات الطماطم في النبات الواحد إذا توفرت لدينا البيانات الآتية:

عدد حبات الطماطم في النبات	2	3	4	5	6	7	8	9
f عدد النباتات	4	5	10	29	30	18	3	1

الحل

أولاً: نقوم بإدخال البيانات للبرنامج بإتباع الخطوات التالية:

١- من نافذة Variable View سوف نقوم بتعريف متغيرين الأول X يمثل عدد حبات الطماطم في النبات والثاني f يمثل عدد النباتات

٢- نقوم بإدخال بياناتها من نافذة Data View

*ex3_3.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

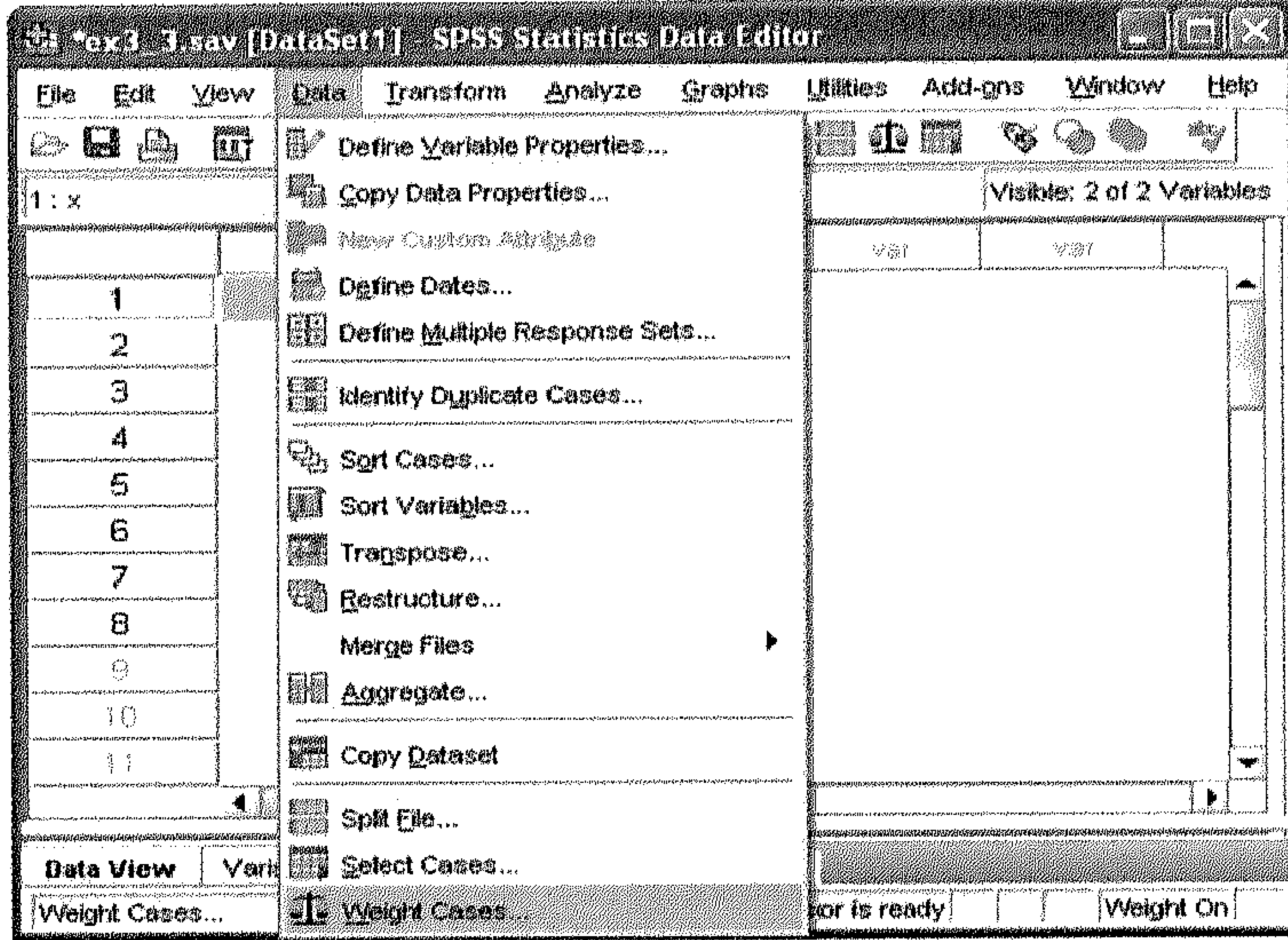
1 : x 2.0 Visible: 2 of 2 Variables

	x	f
1	2.00	4.00
2	3.00	5.00
3	4.00	10.00
4	5.00	29.00
5	6.00	30.00
6	7.00	18.00
7	8.00	3.00
8	9.00	1.00

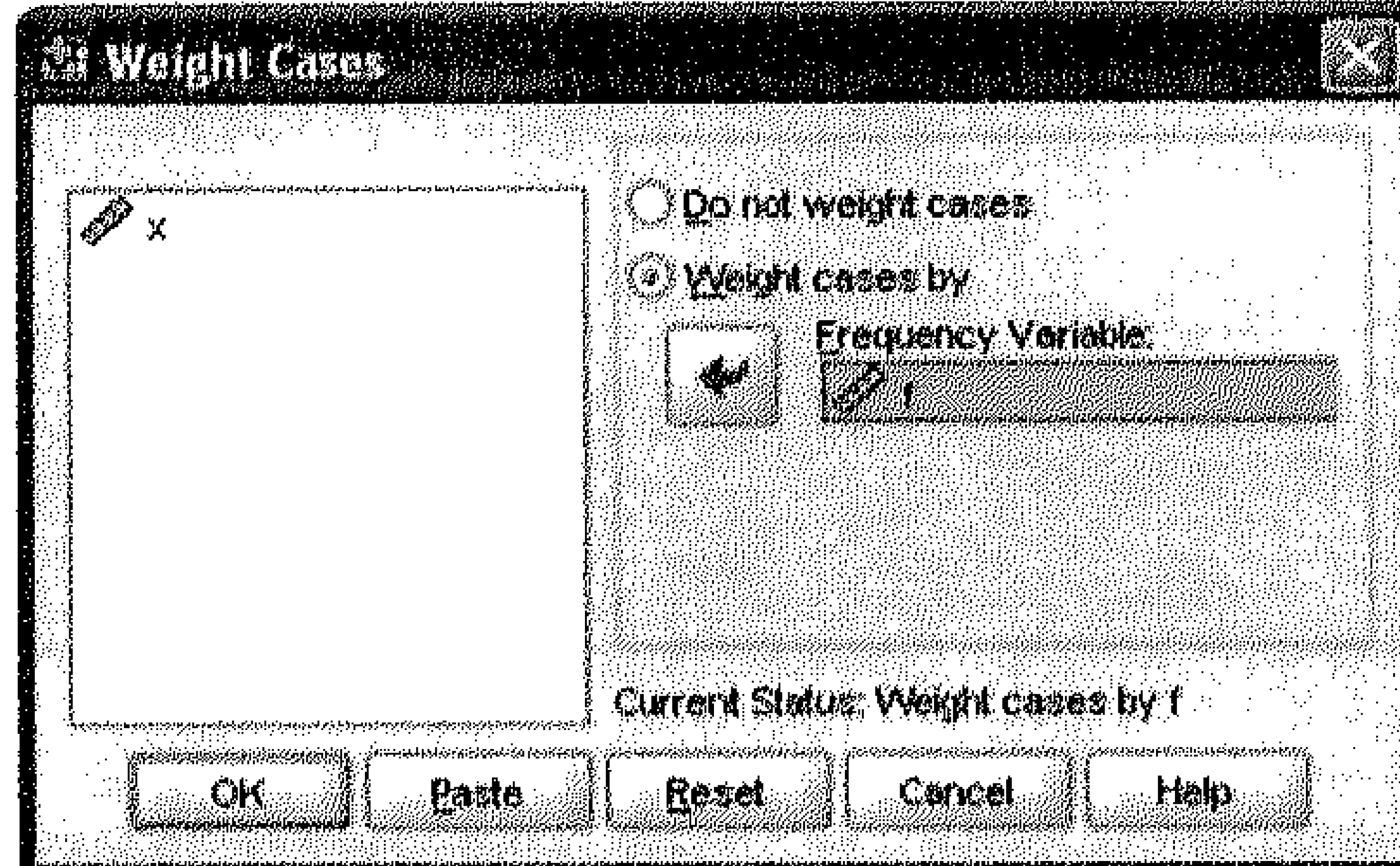
Data View Variable View

SPSS Statistics Processor is ready Weight On

٣- من قائمة Data نختار Weight Cases لتحديد أن f هي تكرار X



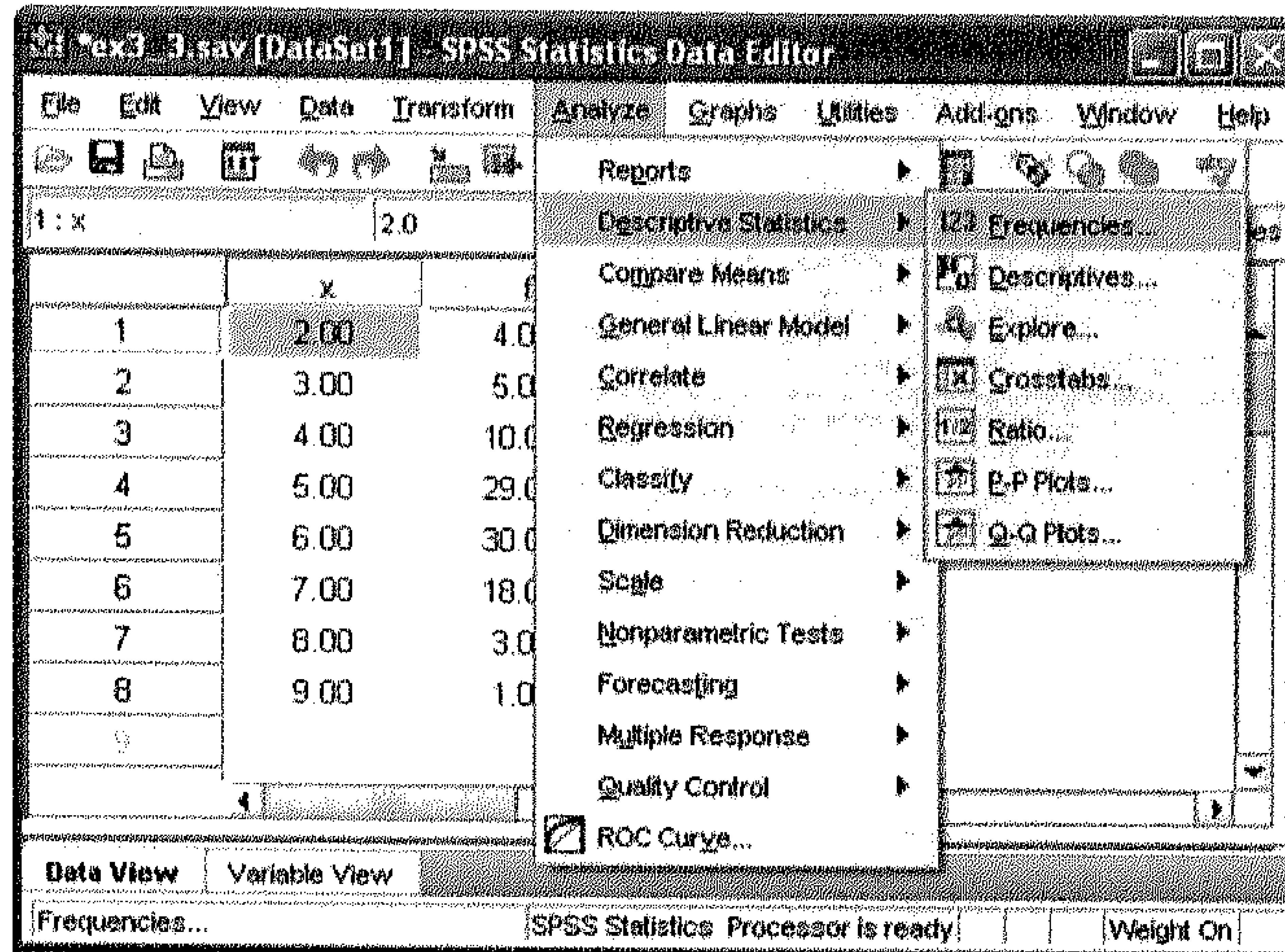
٤- تظهر شاشة بعنوان Weight Cases نجد أن الاختيار Do not weight cases هو المحدد فنقوم باختيار weight cases by ثم ننقل المتغير *f* لشريط Frequency Variable ثم نضغط على Ok



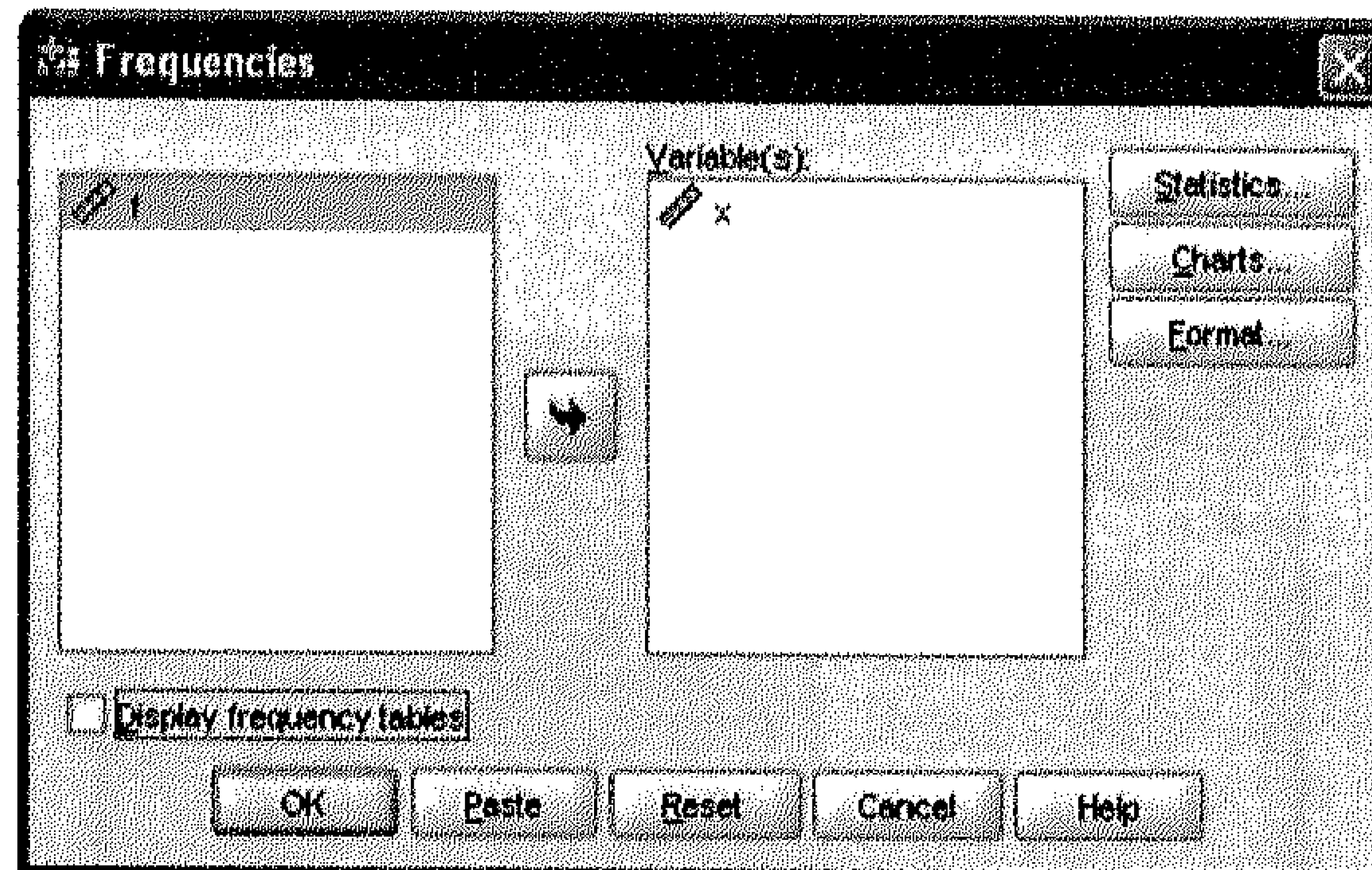
ثانياً: تعيين المقاييس المطلوبة بإتباع الخطوات التالية:

١- نضغط على قائمة Analyze ونختار منها Descriptive Statistics

٢- تظهر قائمة منسدلة فنضغط على Frequencies



٣- تظهر شاشة بعنوان Frequencies ننقل المتغير X لقائمة Variable(s)



٤- نضغط على Statistics فتظهر الشاشة التي تحتوي على المقاييس الإحصائية فمن قائمة Central Tendency نختار Mean, Median, Mode

٥- ومن قائمة Percentile Value نختار Quartiles لتعيين الربيعيات ونختار Percentile(s) لحساب القيم التي قبلها 2% من البيانات وتعني المئين الثاني، وأيضا القيمة التي قبلها 98% من البيانات وتمثل المئين الثامن والتسعين.

Frequencies: Statistics

Percentile Values
Quartiles
Cut points for: 10 equal groups
Percentile(s):
Add Change Remove
2.0
98.0

Central Tendency
Mean
Median
Mode
Sum

Values are group midpoints

Dispersion
Std. deviation Minimum
Variance Maximum
Range S.E. mean

Distribution
Skewness
Kurtosis

Continue Cancel Help

٦- نضغط على Continue فنعود للشاشة السابقة

٧- نضغط على Ok فتظهر النتائج التالية

Statistics

x

N	Valid	100
	Missing	0
Mean		5.4700
Median		6.0000
Mode		6.00
Percentiles	2	2.0000
	25	5.0000
	50	6.0000
	75	6.0000
	98	8.0000

ومن الجدول السابق نجد أن

عدد القيم هو $n = 100$, ، الوسط الحسابي للبيانات $\bar{x} = Mean = 5.47$, ، الوسيط للبيانات $Med = 6$, ، المنوال للبيانات هو $Mod = 6$, ، المئين الثاني هو $P_2 = 2$، المئين الخامس والعشرون الذي يساوى الربيع الأول يساوي $P_{25} = Q_1 = 5$, ، المئين الخمسين الذي يساوي الربيعي الثاني والوسيط يساوي $P_{50} = Q_2 = Med = 6$,، المئين الخامس والسبعون الذي يساوى الربيعي الثالث يساوى $P_{75} = Q_3 = 6$,، المئين الثامن والتسعون يساوي $P_{98} = 8$

تطبيق (٣-٣):

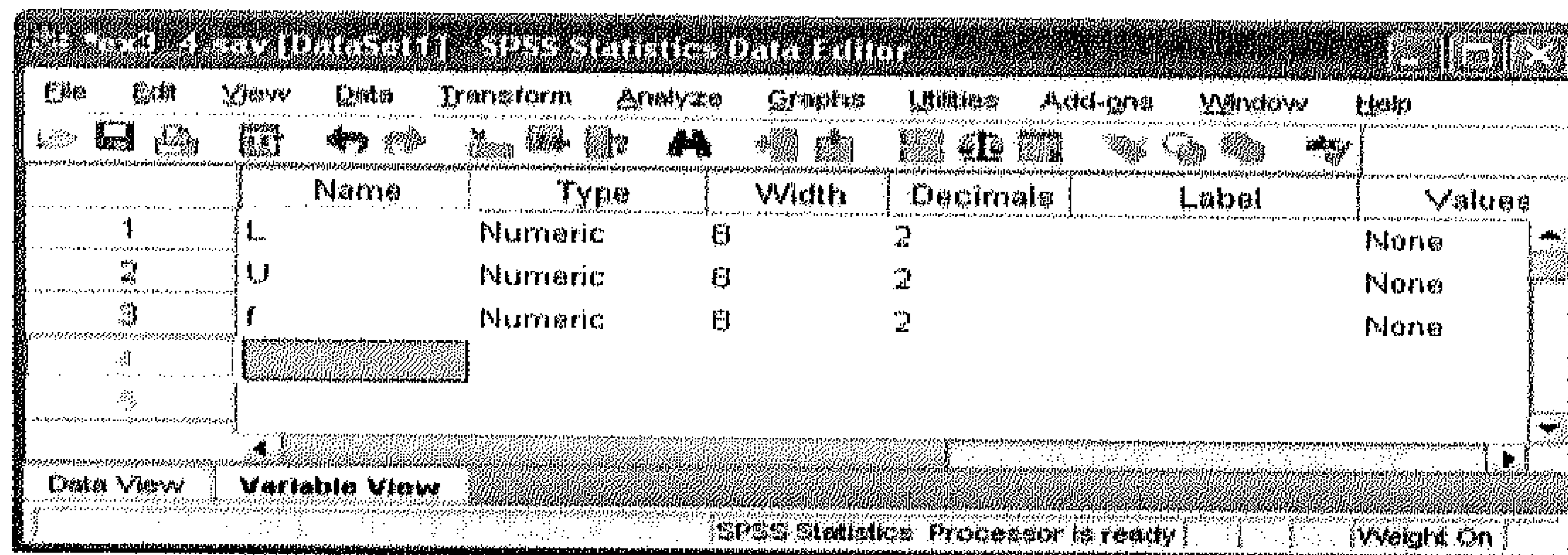
في مثال (٣-٤) أوجد مقاييس النزعة المركزية للتوزيع التكراري الآتي:

الفئات	35–	40–	45–	50–	55–	60–
التكرارات	7	18	23	24	16	12

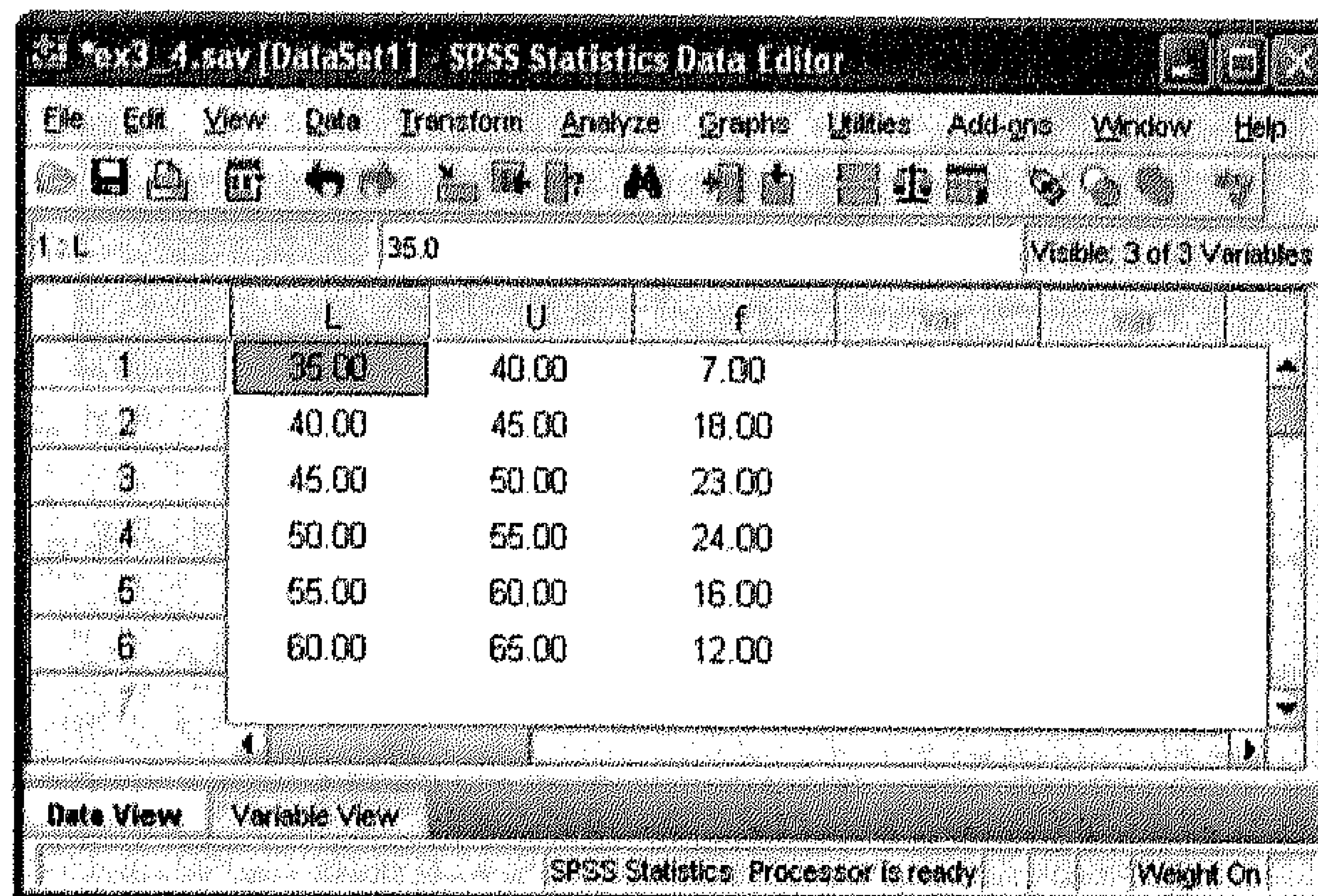
الحل

أولاً: إدخال البيانات بإتباع الخطوات التالية:

١– سوف نقوم بتعريف ثلاثة متغيرات هي L, U, f وهي تمثل الحد الأدنى للفترات والحد الأعلى والتكرار المناظر للفترات من نافذة Variable View

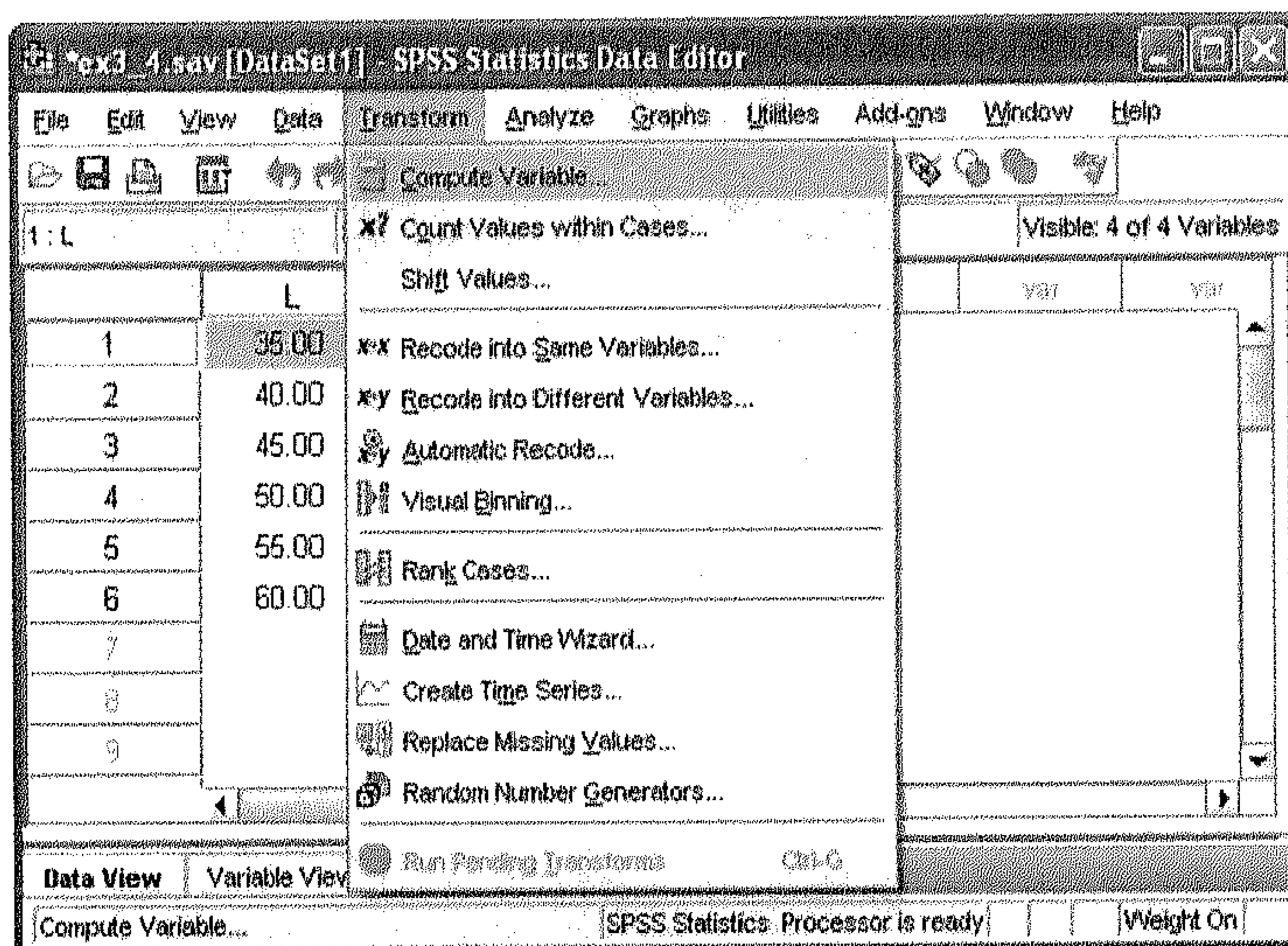


٢– وسنقوم بإدخال البيانات من محرر البيانات (Data View)

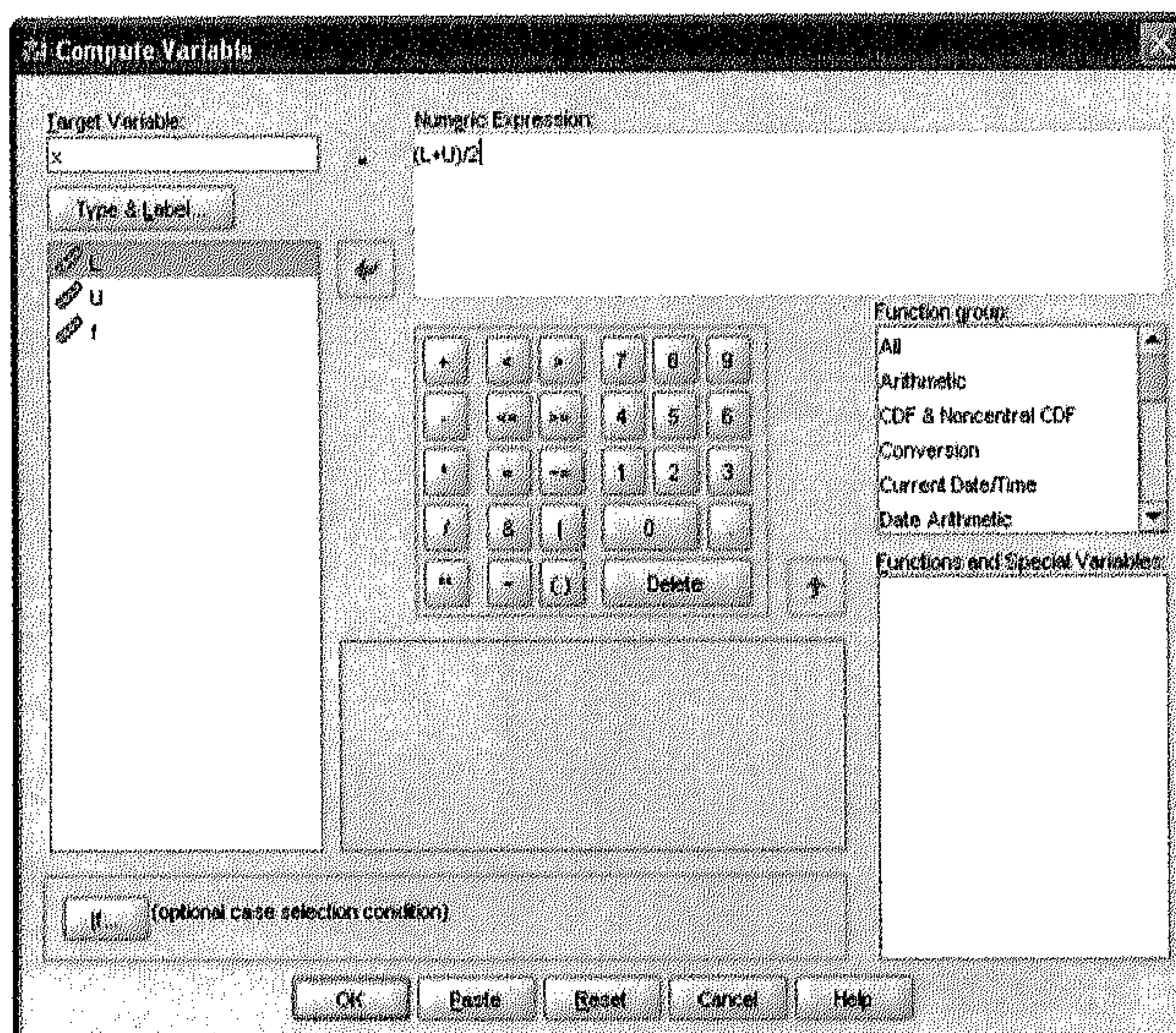


ثانياً: تعيين مركز الفترات بإتباع الخطوات التالية:

١- من قائمة Transform نختار Compute Variable



٢- تظهر شاشة جديدة بعنوان Compute Variable نكتب في خانة Target Variable اسم المتغير الجديد وهو x والذي يمثل مركز الفترات



٣- في قائمة Numeric Expression نكتب العلاقة التي ستستخدم في حساب قيم المتغير x وهو ناتج من قسمة مجموع بداية الفترات مع نهايتها على 2

٤- نضغط على Ok نعود للملف وقد أضاف متغيراً جديداً X يحتوي على مركز الفترات

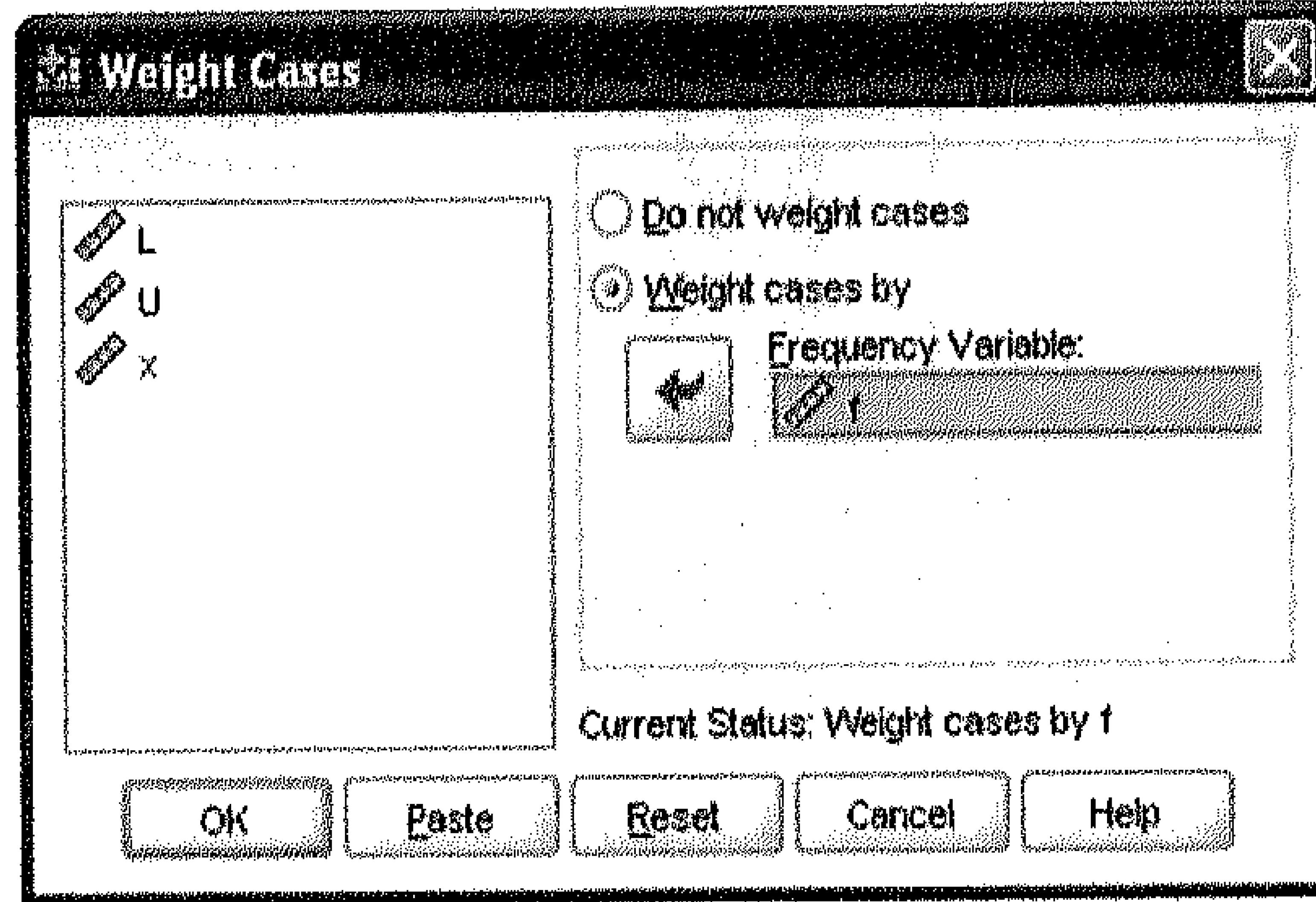
	L	U	f	x
1	35.00	40.00	7.00	37.50
2	40.00	45.00	18.00	42.50
3	45.00	50.00	23.00	47.50
4	50.00	55.00	24.00	52.50
5	55.00	60.00	16.00	57.50
6	60.00	65.00	12.00	62.50

ثالثاً: يجب تحديد أن f هي تكرار مناظر لمركز الفترات

١- من قائمة Data نختار Weight Cases

٢- تظهر شاشة جديد بعنوان Weight Cases نختار المتغير x ثم نختار Weight cases by

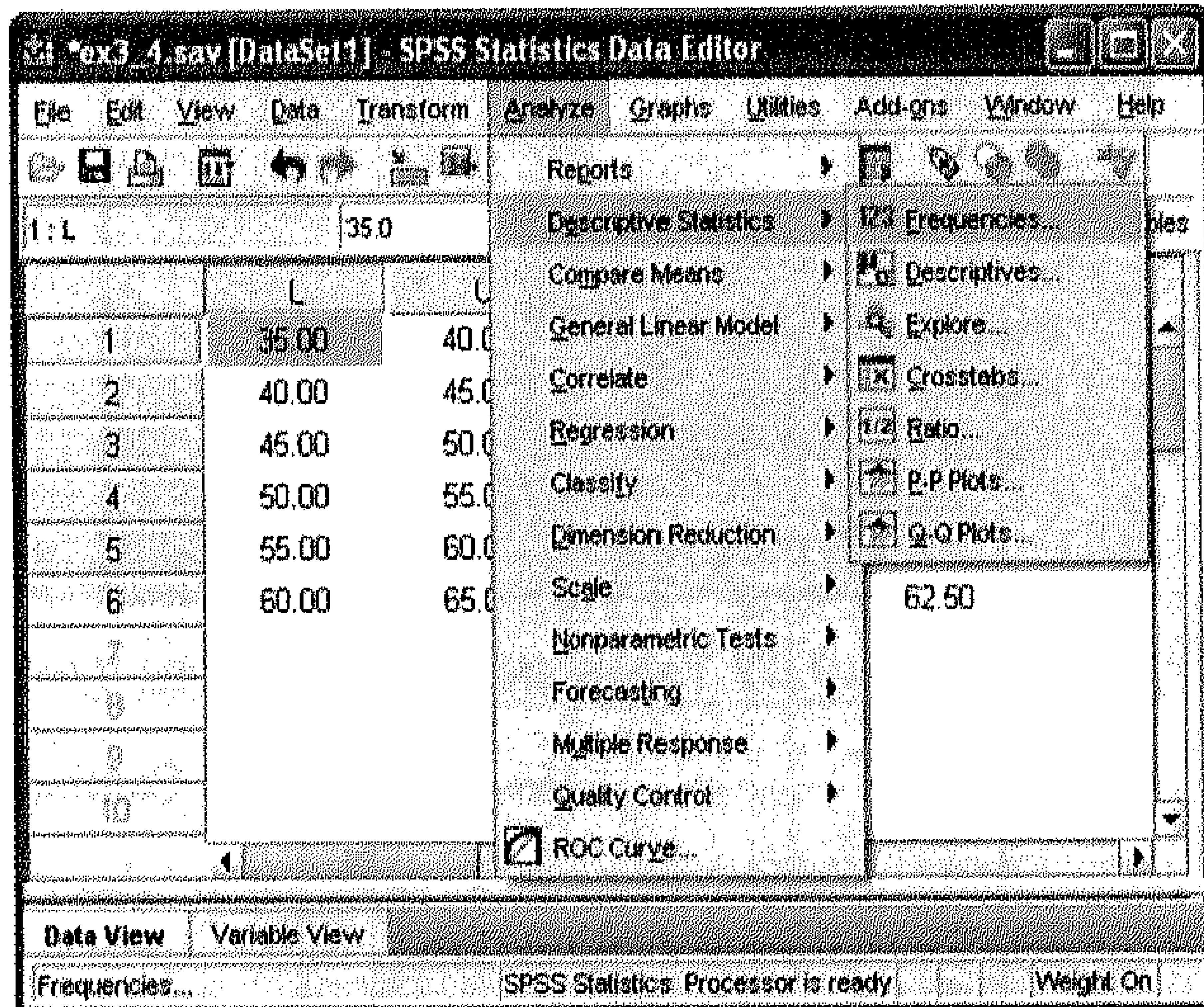
٣- ننقل المتغير f لشريط Frequency Variable



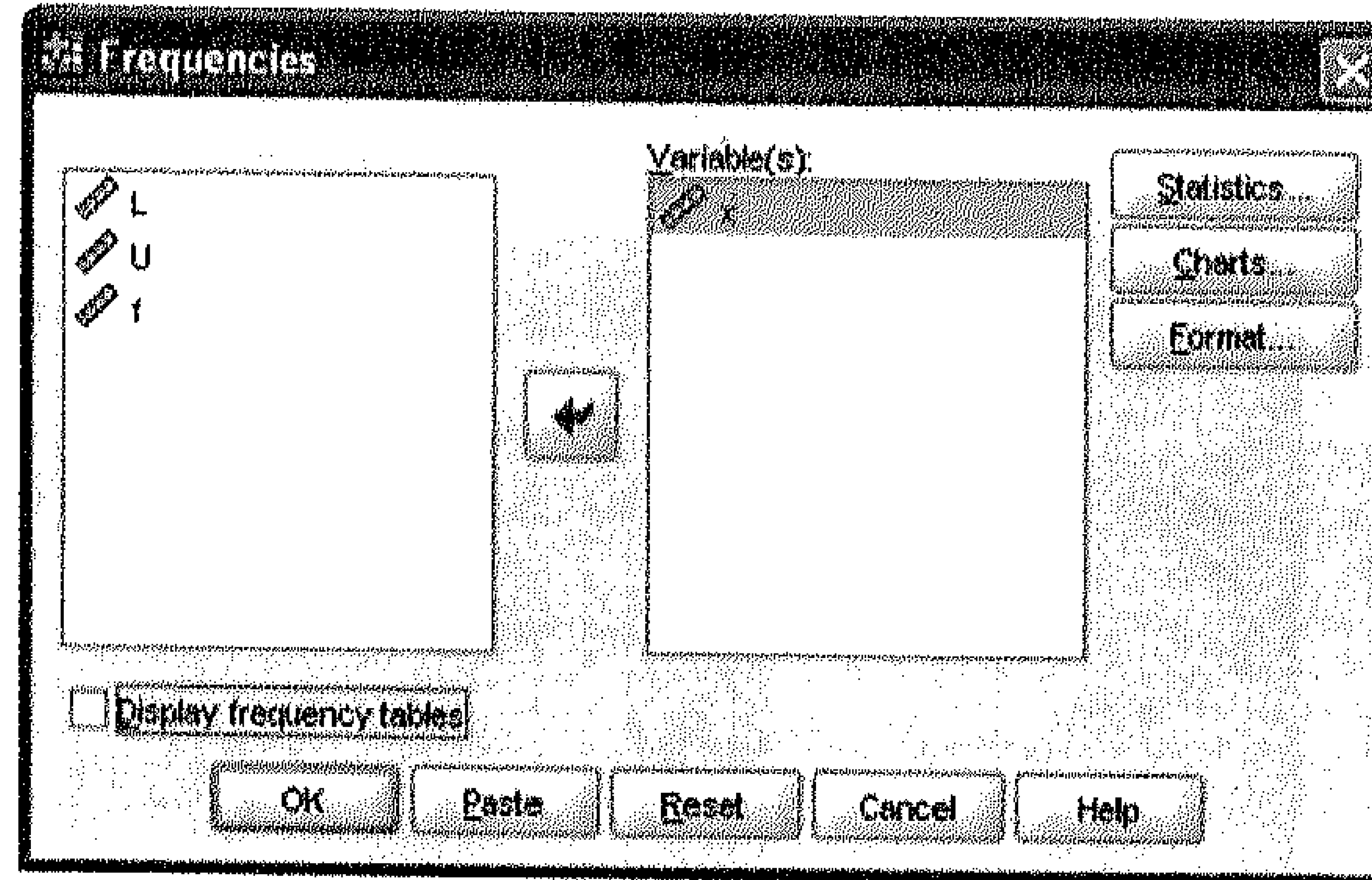
٤- ثم نضغط على Ok فنعود لملف البيانات

رابعاً: تعيين المقاييس المطلوبة بإتباع الخطوات التالية:

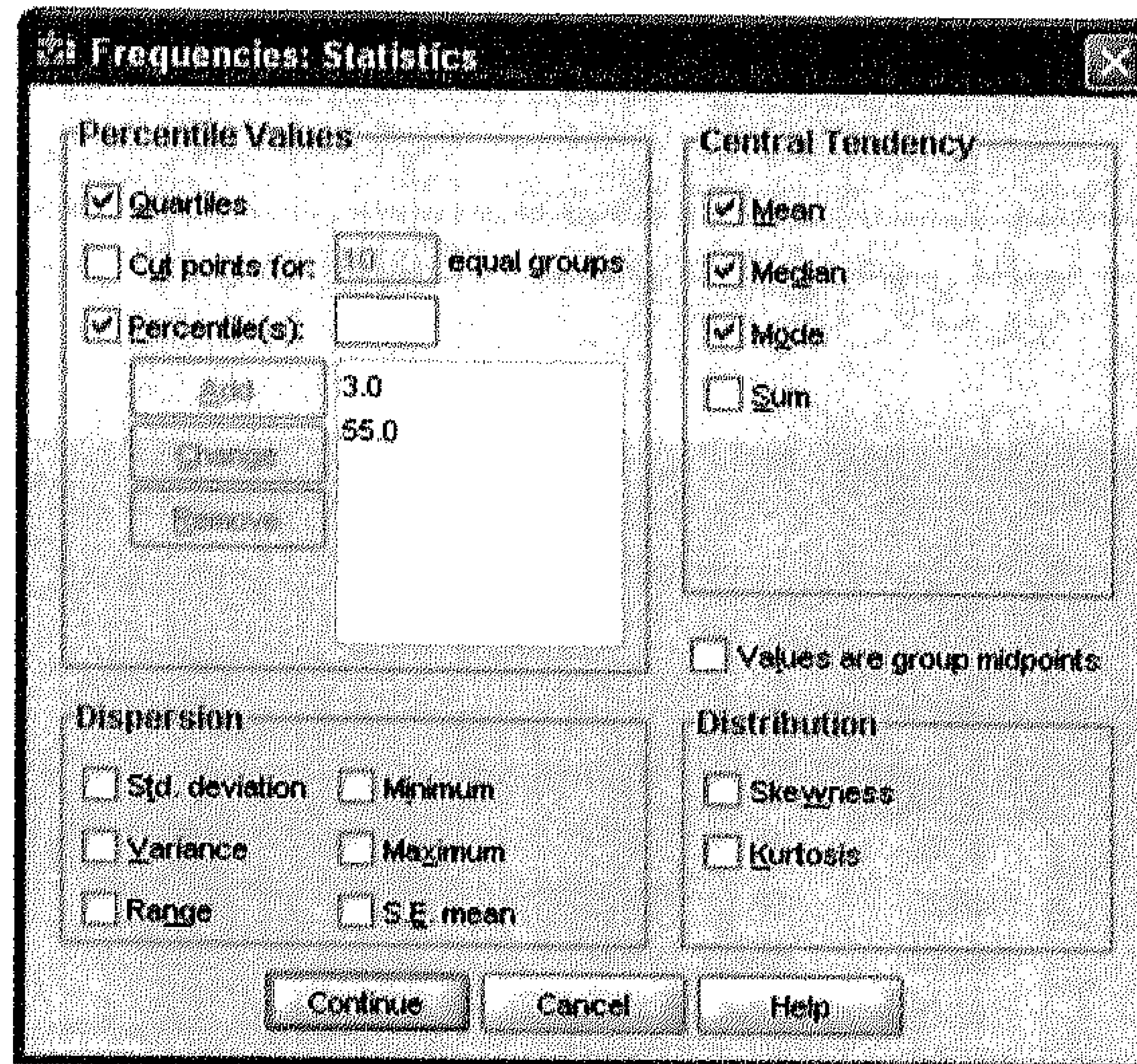
١- من قائمة Analyze نختار الأمر Statistics Descriptive فتظهر قائمة منسدلة نختار منها الأمر Frequencies



٢- ننقل المتغير X لقائمة Variable(s) ثم نضغط على الأمر Statistics



٣- نختار كلاً من Mode ,Median ,Mean من قائمة Central Tendency ونختار Quartiles وأيضاً Percentile(s) من قائمة Percentiles Values والتي نحدد فيها المئين الثالث وأيضا الخامس والخمسين.



لاحظ إننا قد اخترنا Values are group midpoints لأن قيم X هي مركز الفترات

٤- نضغط على Continue فنعود للشاشة السابقة

٥- نضغط على Ok فتظهر النتائج التالية:

Statistics

x

N	Valid	100
	Missing	0
Mean		50.5000
Median		50.3723[a]
Mode		52.50
Percentiles	3	.[b]
	25	44.6951
	50	50.3723
	55	51.4362
	75	56.2500

a. Calculated from grouped data.

b. The lower bound of the first interval or the upper bound of the last interval is not known. Some percentiles are undefined.

c. Percentiles are calculated from grouped data.

من الجدول السابق فإن:

عدد القيم يساوي $n = 100$ ، الوسط الحسابي للبيانات يساوى $\bar{x} = 50.5$ ، الوسيط للبيانات يساوي $Med = 50.3723$ ، المنوال للبيانات يساوي $Mod = 52.5$ ،المئين الخامس والعشرون ويساوي الربيع الأول هو $P_{25} = Q_1 = 44.6951$ ، المئين الخمسون ويساوي الربيع الثاني وأيضا يساوي الوسيط هو $P_{50} = Q_2 = Med = 50.3723$ ، المئين الخامس والخمسون يساوي $P_{55} = 51.43625$ ، المئين الخامس والسبعون الذي يساوي الربيع الثالث يساوي $P_{75} = Q_3 = 56.25$

تطبيق (٣-٤):

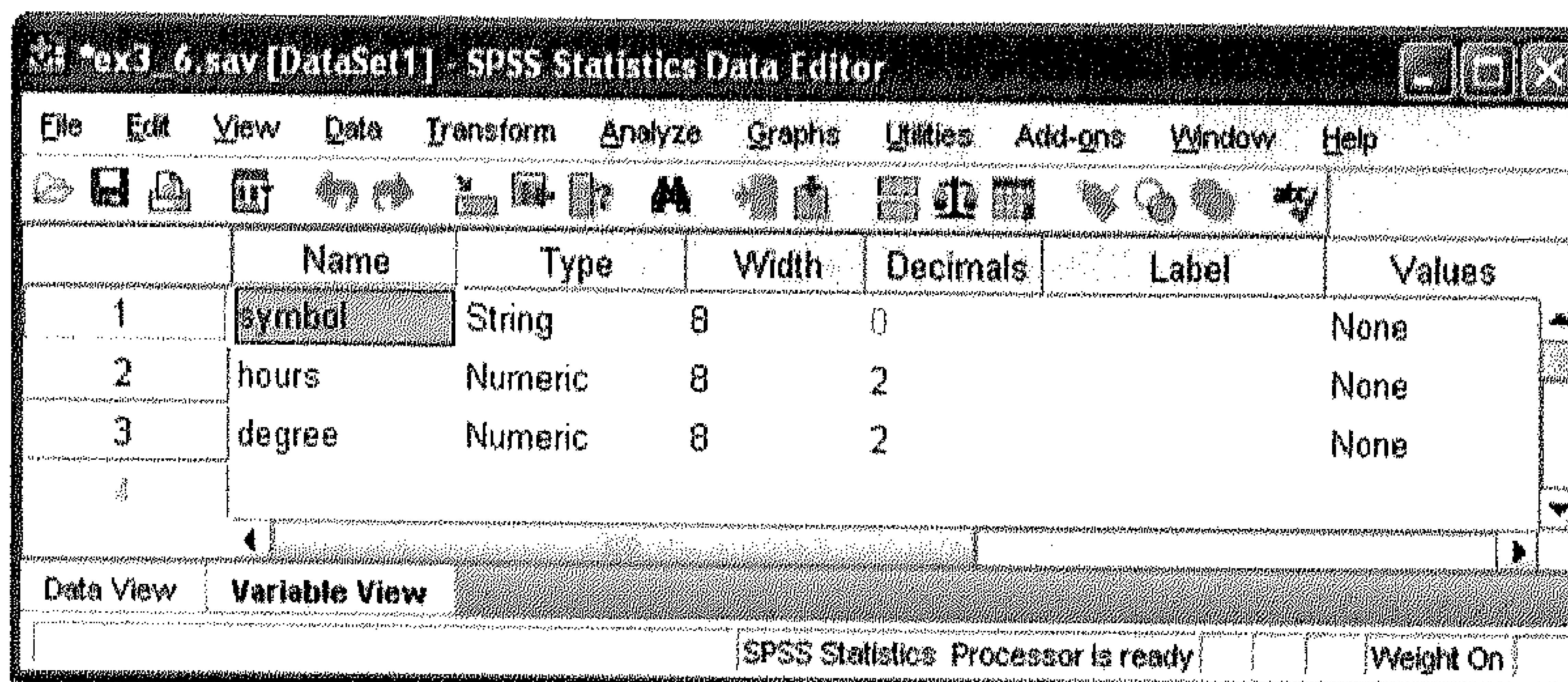
من مثال (٣-٦): أوجد المعدل لأحد طلاب المستوى الثاني بكلية العلوم والدراسات الإنسانية بالخرج إذا كانت درجاته كالتالي:

رمز المقرر	عدد ساعات المقرر	الدرجات
١٠٨ أحص	3	85
١٠٨ حيا	3	72
١٠٨ فيز	3	77
١٠٨ ريض	3	69
١٠٨ نجم	2	79
١٢٣ نجم	3	82

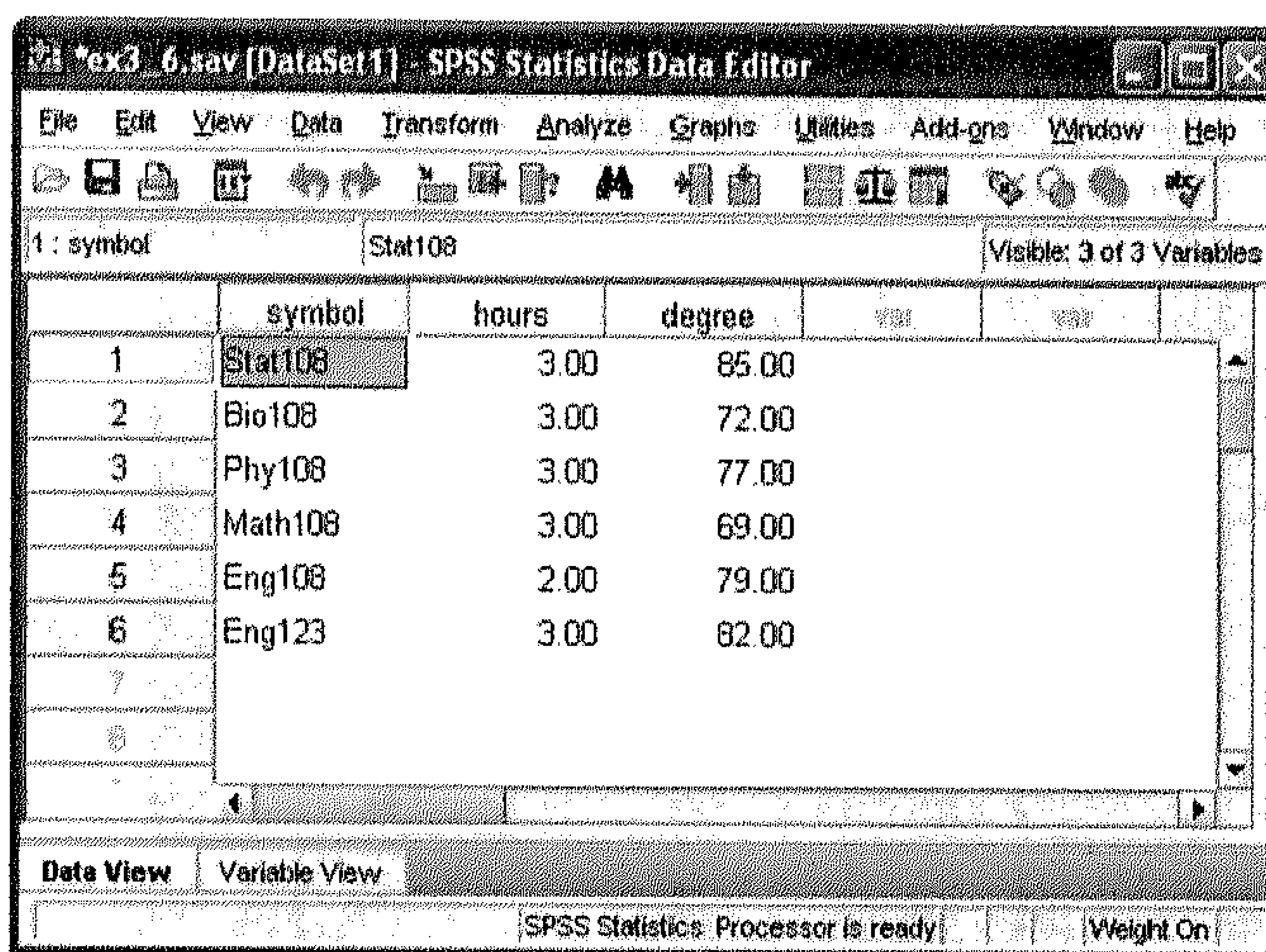
الحل:

أولاً: إدخال البيانات للبرنامج بإتباع الخطوات التالية:

١- من نافذة Variable View سوف نقوم بتعريف ثلاثة متغيرات symbol, hours, degree حيث المتغيرين hours, degree متغيرات عددية لكن المتغير symbol من النوع الحرفي

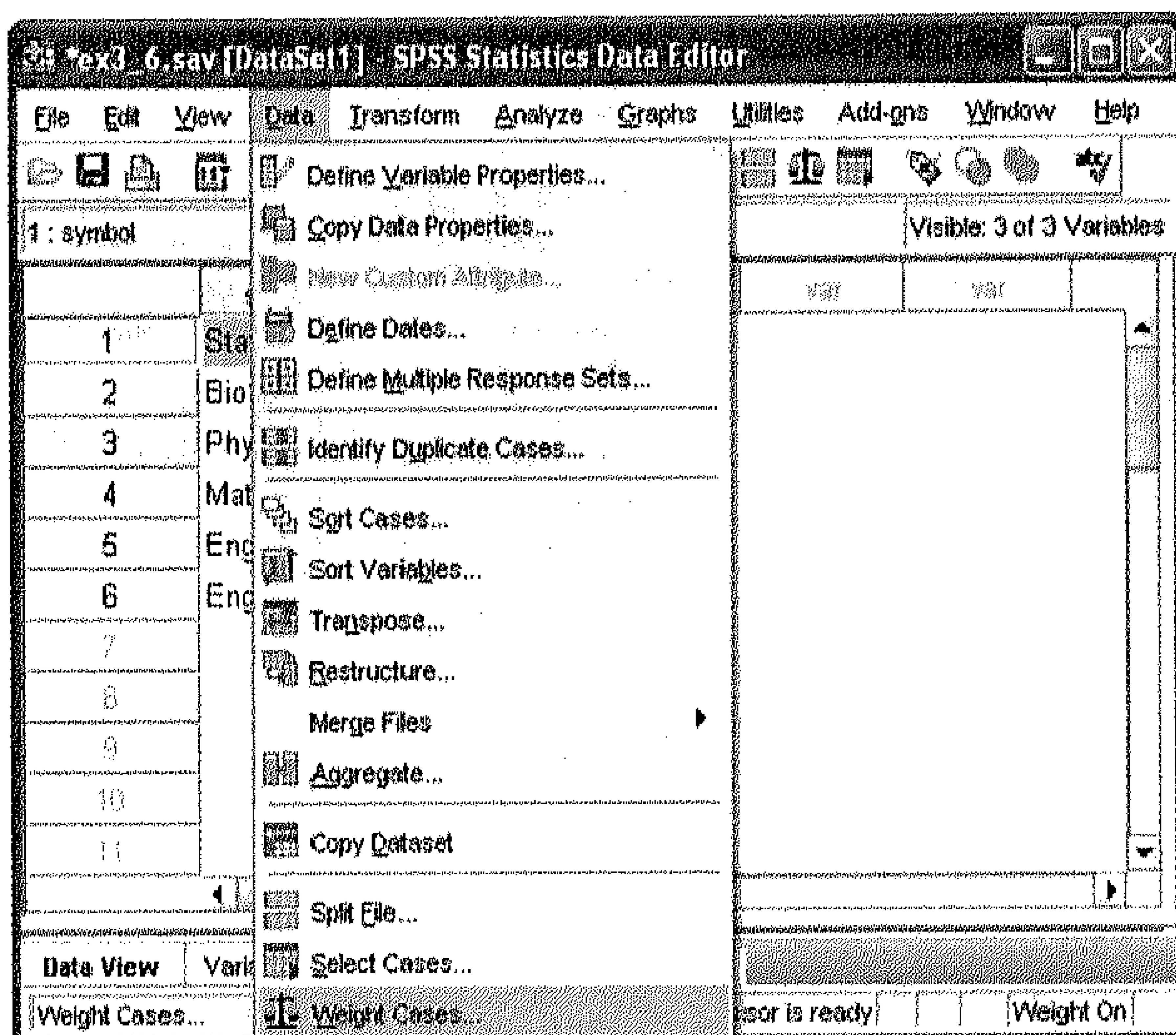


٢- نقوم بإدخال قيم المتغيرات من نافذة Data View



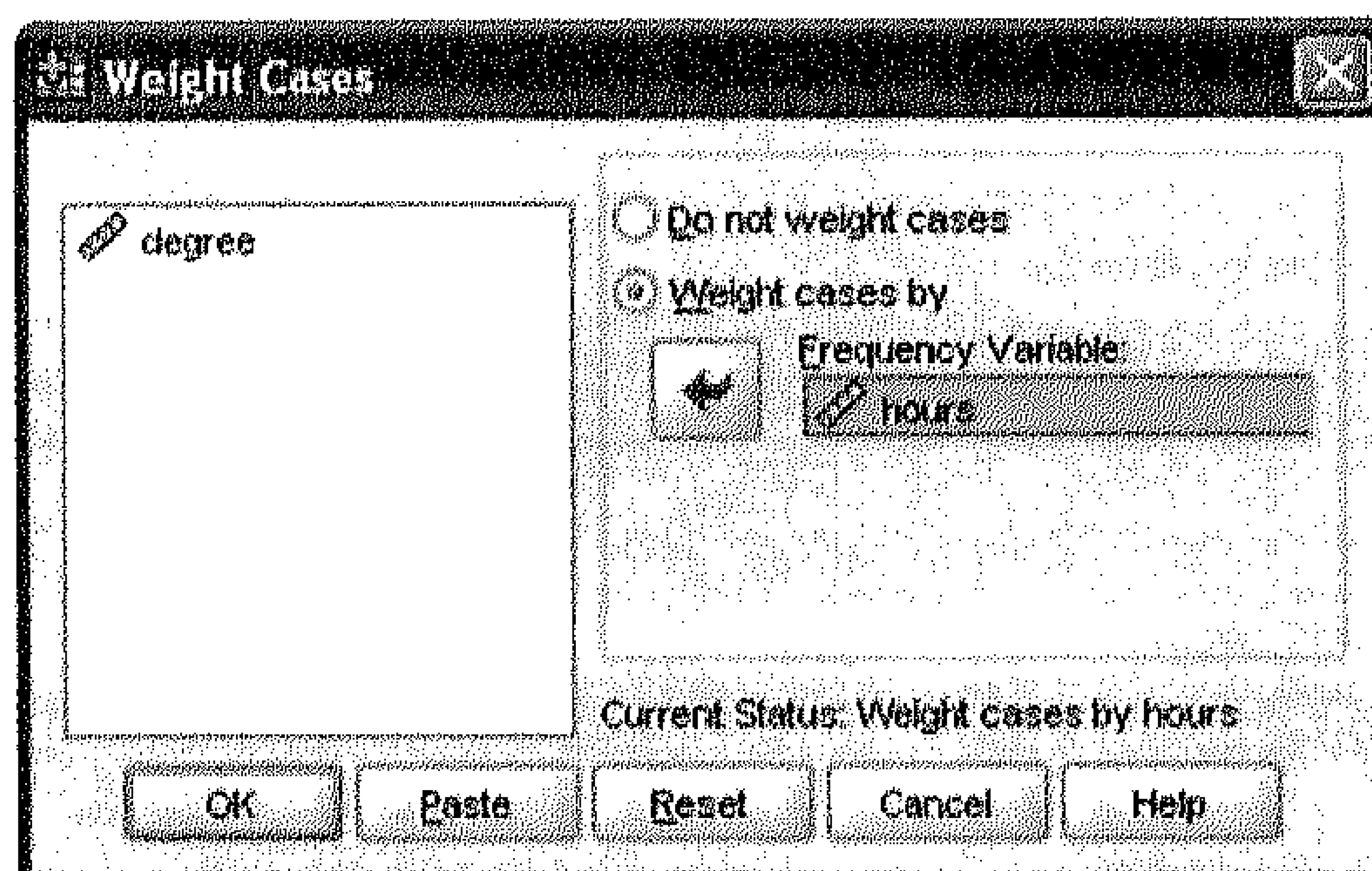
ثانياً: نقوم بتعريف الساعات hours على أنها أوزان للدرجات degree باتباع الخطوات التالية:

١- من قائمة Data نختار weight cases



٢- تظهر نافذة بعنوان Weight Cases فنختار Weight cases by

٣- ثم ننقل المتغير hours لشريط Frequency Variable

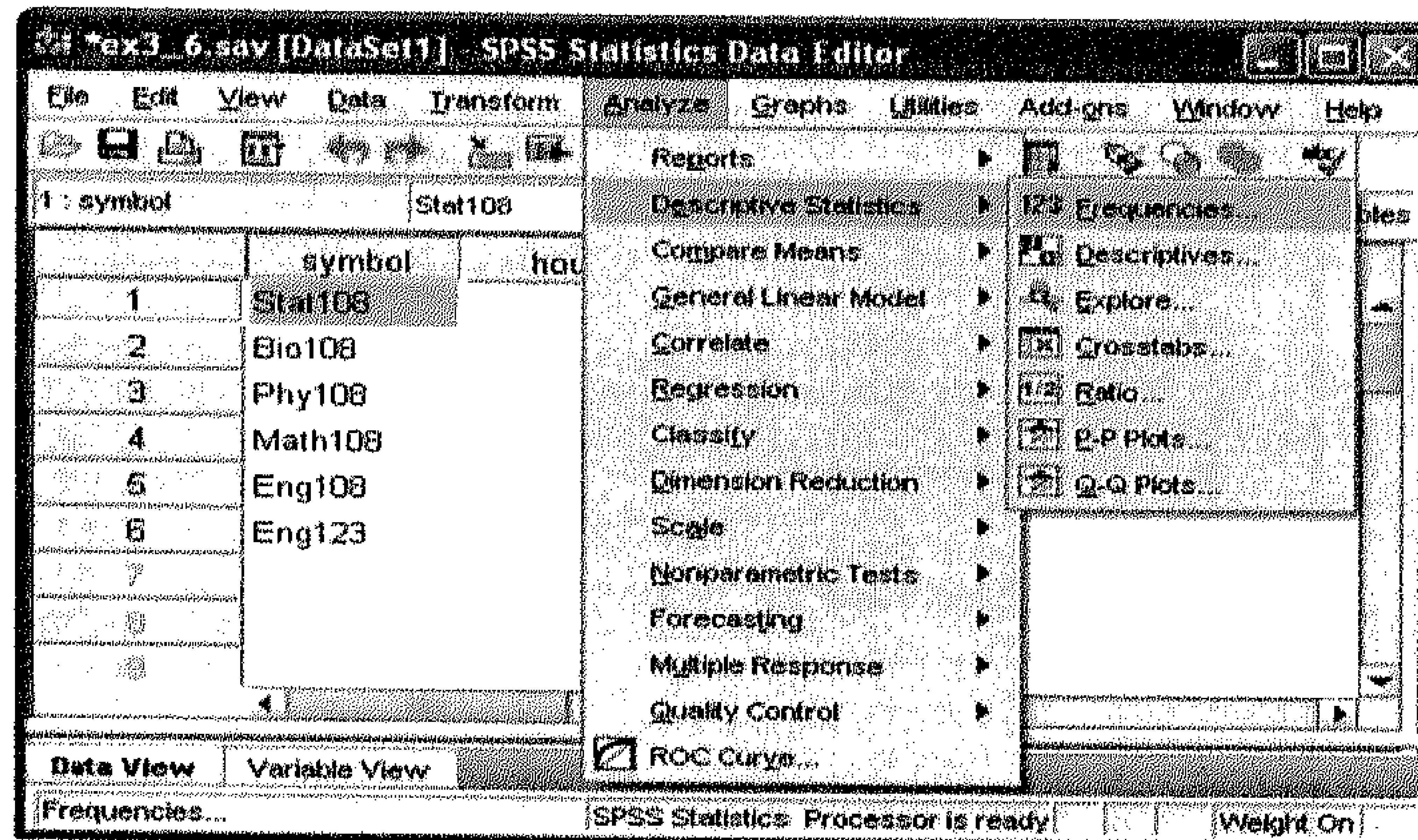


٤- نضغط على Ok فنعود لملف البيانات دون حدوث تغيرات ظاهرية في ملف البيانات.

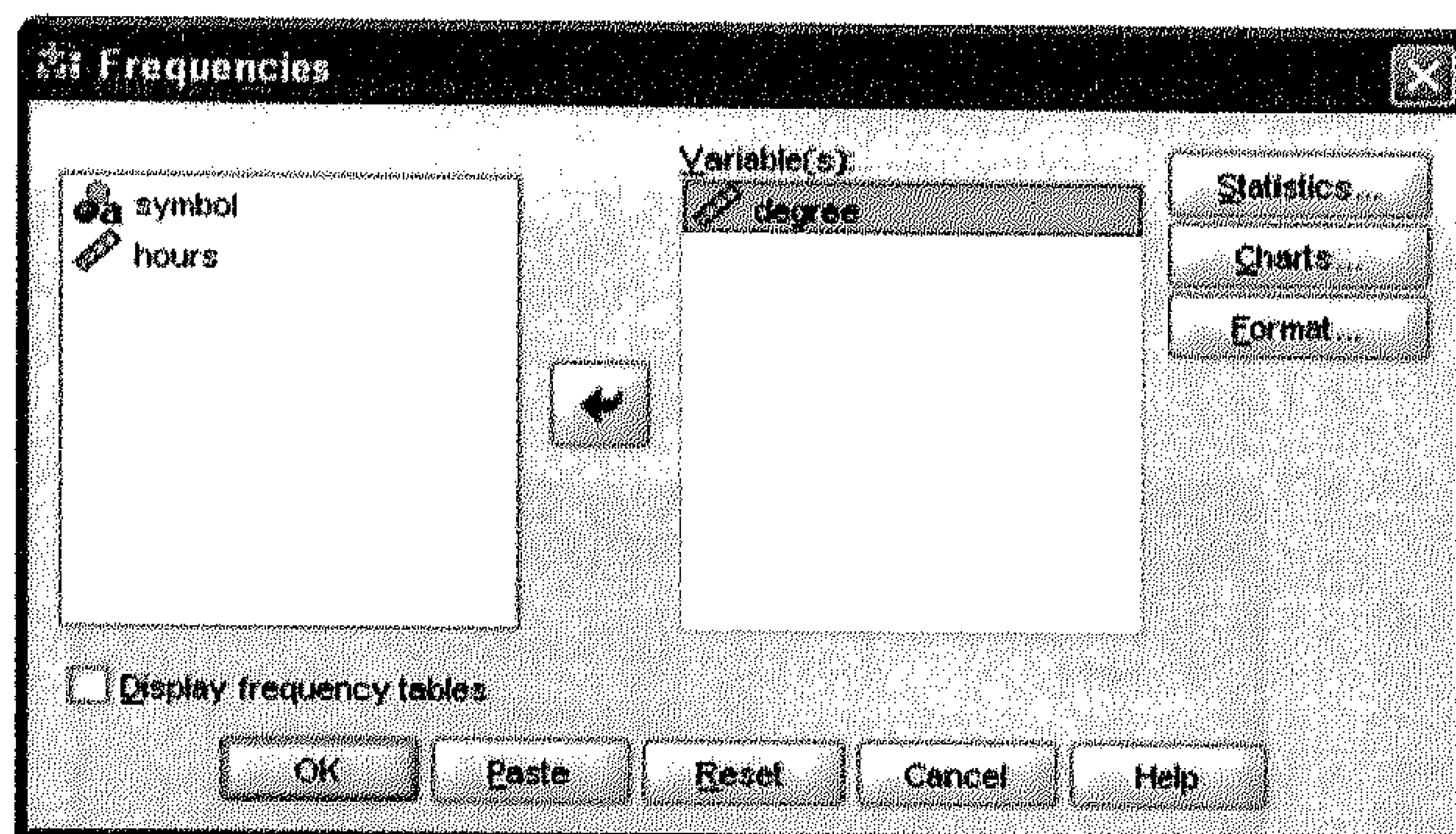
ثالثاً تعيين المقاييس المطلوبة باتباع الخطوات التالية:

١- من قائمة Analyze نختار الأمر Descriptive Statistics

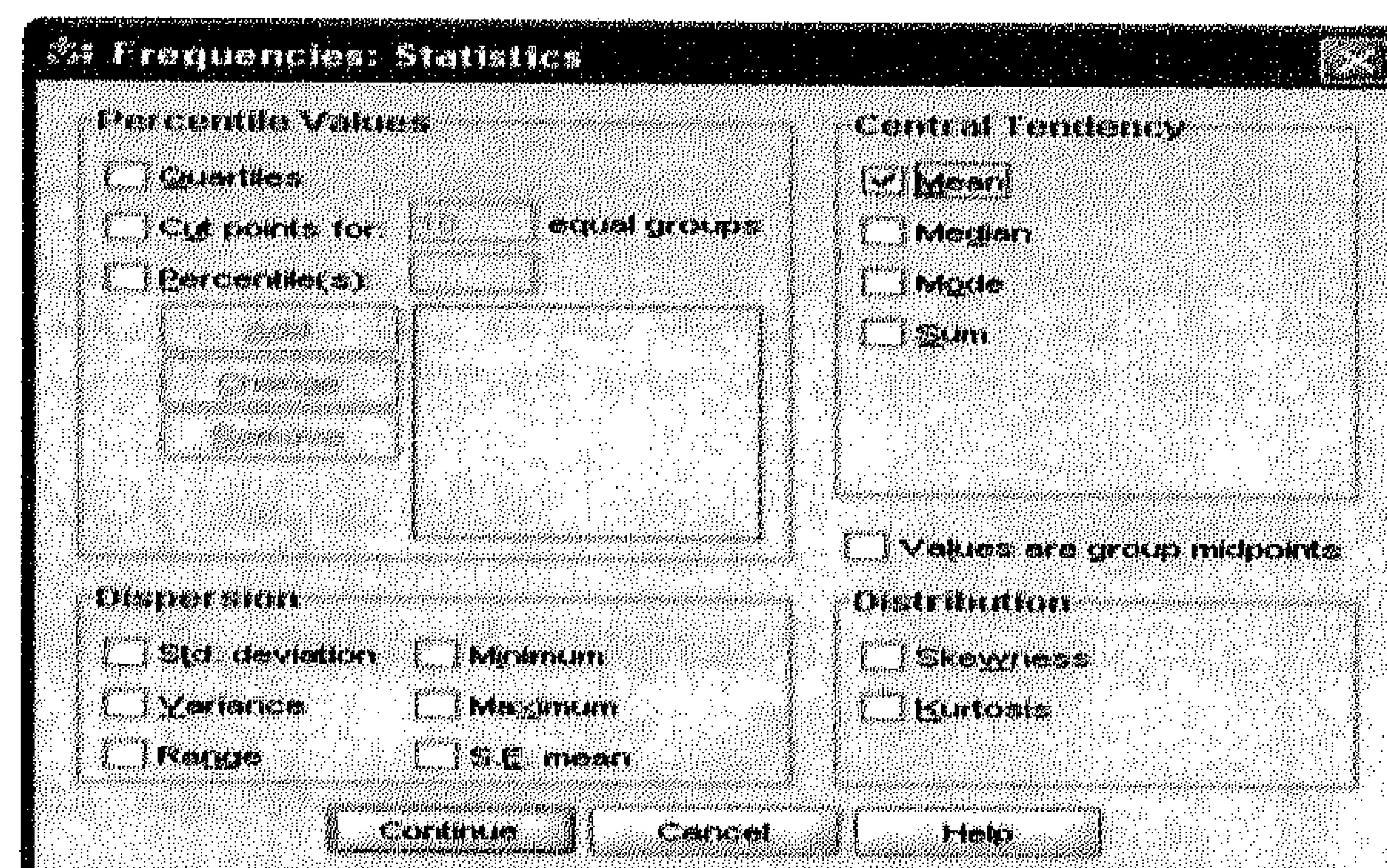
٢- تظهر قائمة منسدلة نختار منها الأمر Frequencies



٥- ننقل المتغير degree لقائمة Variable(s) ثم نضغط على الأمر Statistics



٦- نختار Mean من قائمة Central Tendency



٧- نضغط على Continue فنعود للشاشة السابقة ثم نضغط على Ok فتظهر النتائج التالية:

Statistics

degree

N	Valid	17
	Missing	0
Mean		77.2353

مـن الجـدول السـابق نجـد إن عـدد القـيم المتاحـة هـي $n = 17$، الوسـط الحسـابي(المرجح) للبيانـات يسـاوى $\bar{x} = Mean = 77.2353$

تطبيق (٣-٥):

في مثال (٣-٢٠) الجدول الآتي يبين توزيع 98 نباتاً حسب أطوالها بالسنتيمتر

فئات الطول	10–	15–	18–	20–	22–	25–30
عدد النباتات	15	24	22	20	12	5

المطلوب إيجاد مقاييس النزعة المركزية لطول النبات.

الحل

أولاً: نقوم بإدخال البيانات للبرنامج باتباع الخطوات التالية:

١- من نافذة Variable View نعرف ثلاثة متغيرات L, U, f الحد الأدنى، الحد الأعلى والتكرار المقابل للفترات

٢- من نافذة Data View نقوم بإدخال البيانات

*ex3_19.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : L 10.0 Visible: 3 of 3 Variables

	L	U	f
1	10.00	15.00	15.00
2	15.00	18.00	24.00
3	18.00	20.00	22.00
4	20.00	22.00	20.00
5	22.00	25.00	12.00
6	25.00	30.00	5.00

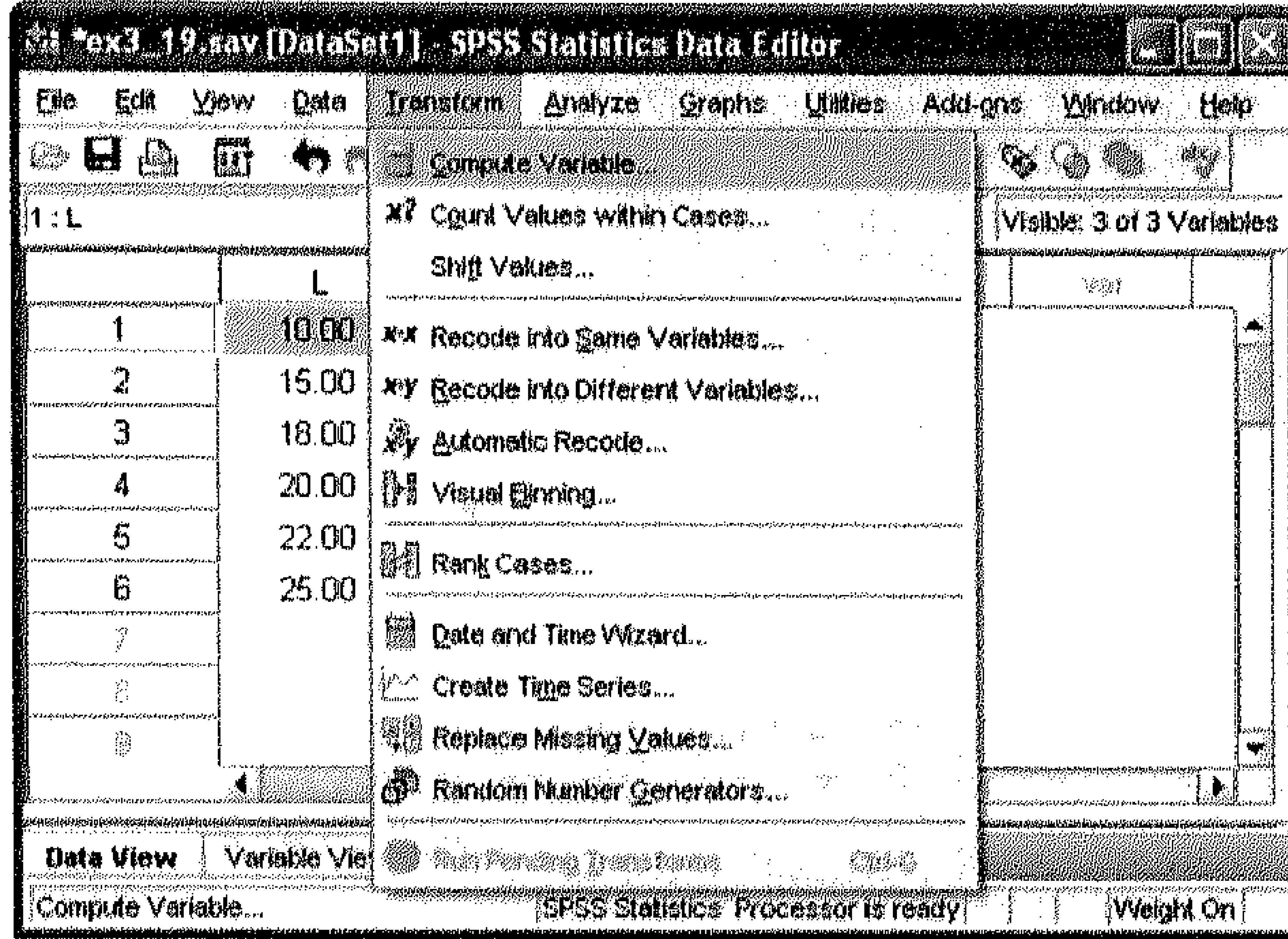
Data View Variable View

SPSS Statistics Processor is ready Weight On

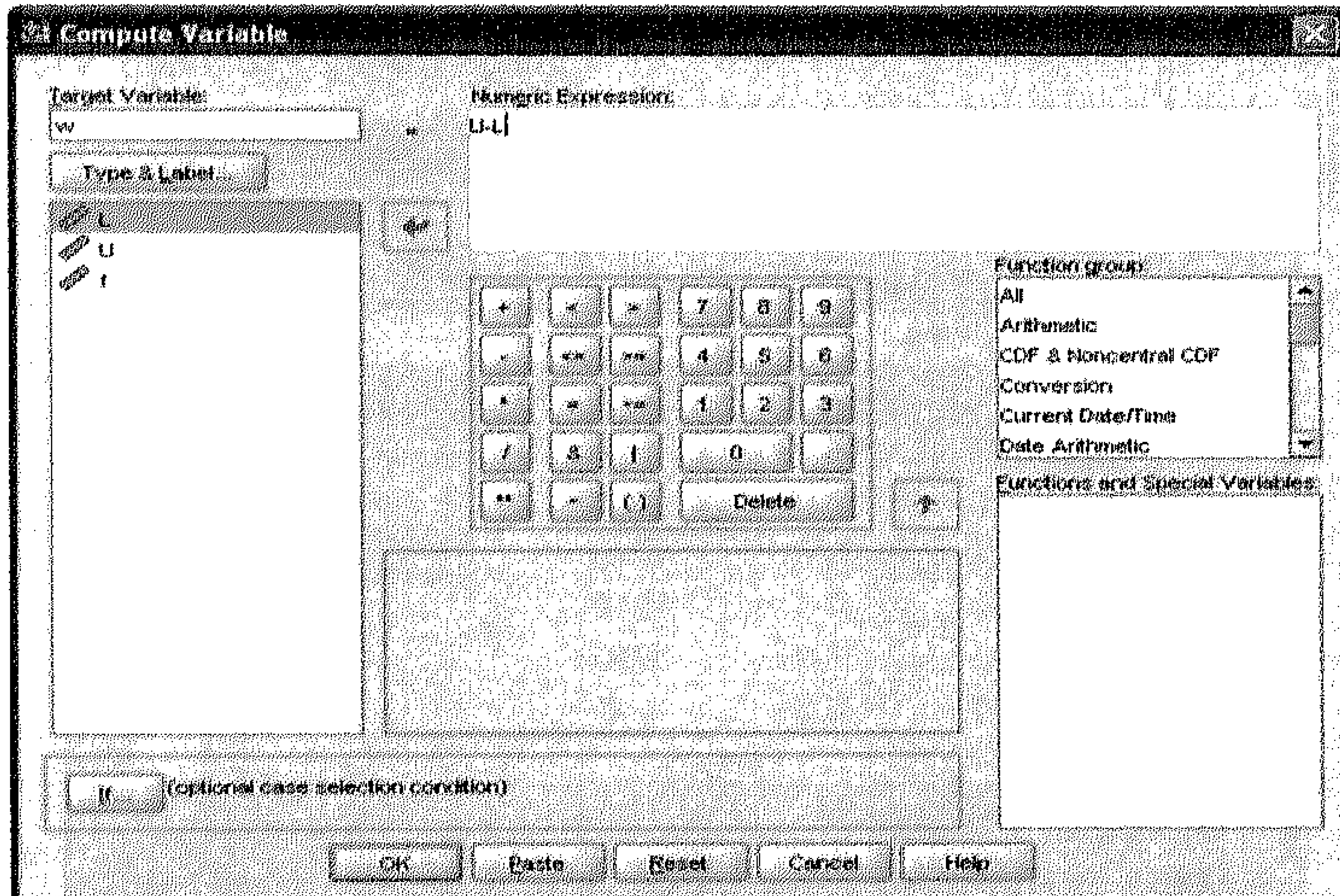
ثانياً: تعيين أطوال الفترات بإتباع الخطوات التالية:

نلاحظ أن أطوال الفترات غير متساو لذا يجب تعيين أطوال الفترات

١- من قائمة Transform نختار Compute Variable



٢- تظهر شاشة جديدة بعنوان Compute Variable نكتب اسم المتغير الجديد W ويمثل طول الفترات بالفرق بين نهاية وبداية الفترة الفعلية $w = U - L$ تكتب في Numeric Expression



٣- نضغط على Ok فنعود للملف وقد أضاف متغيراً جديداً بعنوان *w* ونلاحظ من القيم أنها غير متساوية

*ex3_19.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : L 10.0 Visible: 4 of 4 Variables

	L	U	f	w	var
1	10.00	15.00	15.00	5.00	
2	15.00	18.00	24.00	3.00	
3	18.00	20.00	22.00	2.00	
4	20.00	22.00	20.00	2.00	
5	22.00	25.00	12.00	3.00	
6	25.00	30.00	5.00	5.00	
7					
8					

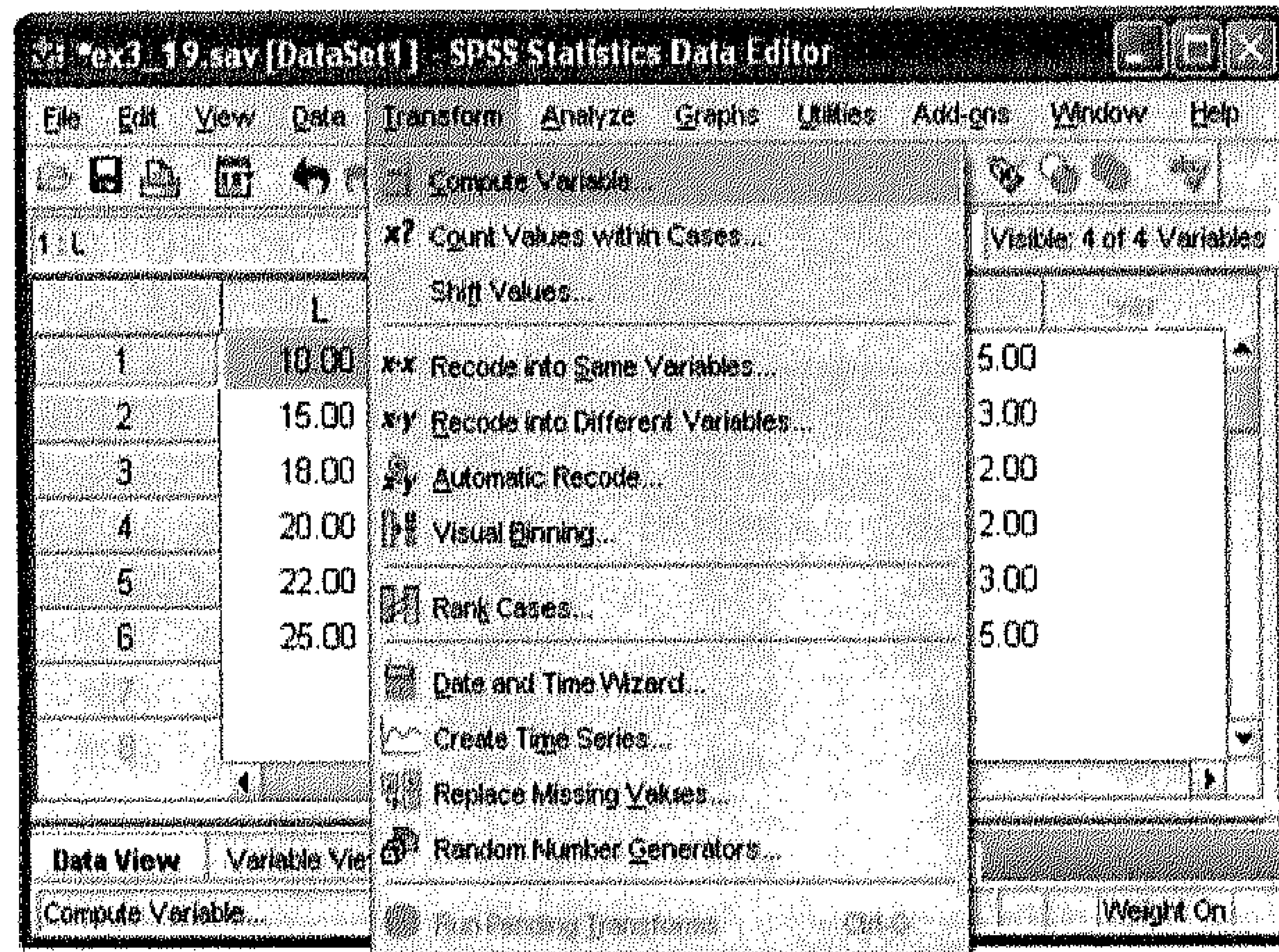
Data View Variable View

SPSS Statistics Processor is ready Weight On

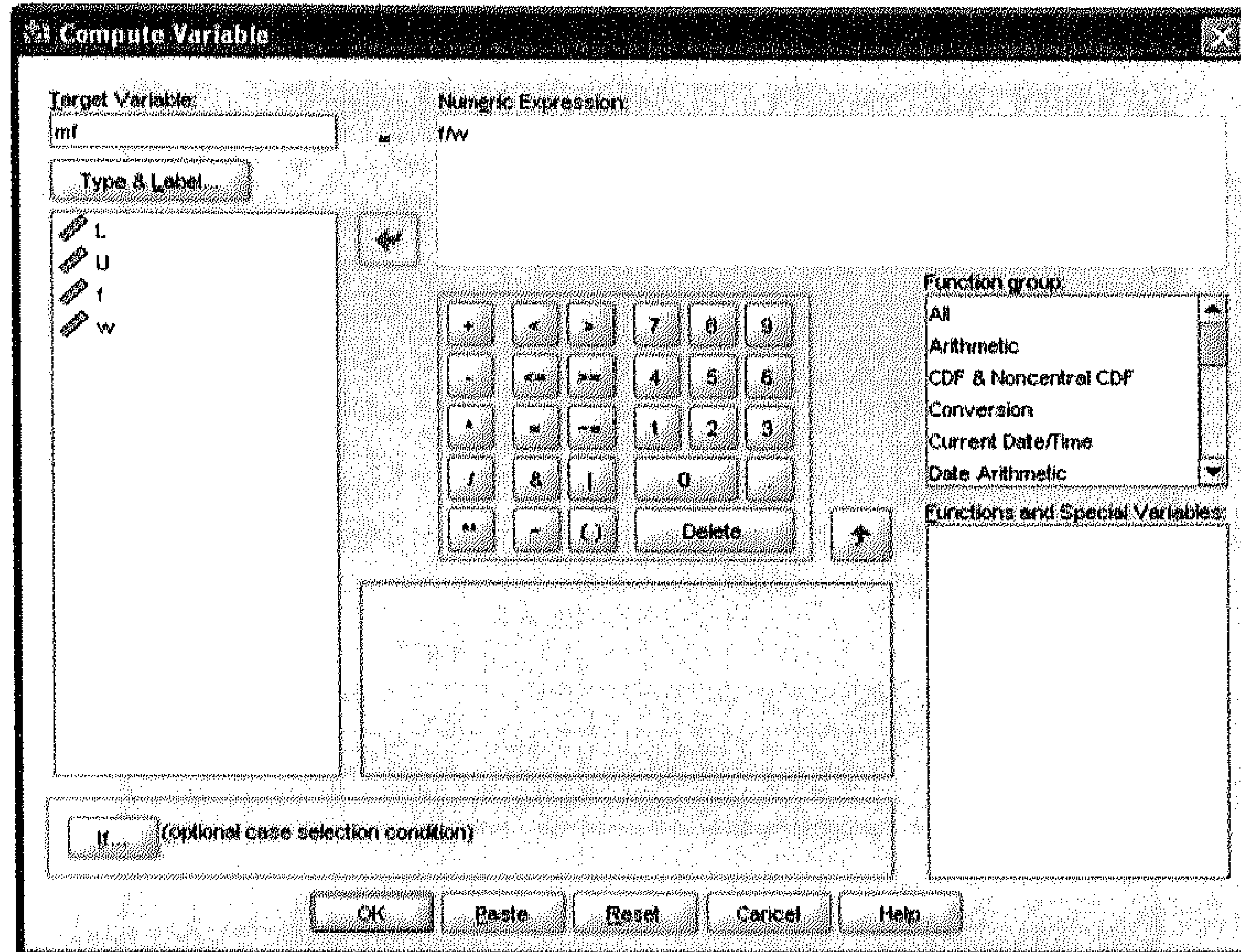
ثالثاً: تعيين التكرار المعدل باتباع الخطوات التالية:

نتيجة لأن الفترات غير متساوية الطول فيجب استخدام التكرار المعدل بدلاً من التكرار

١- من قائمة Transform نختار Compute Variable



٢- تظهر شاشة جديدة بعنوان Compute Variable



٣- تكتب اسم المتغير الجديد mf ويمثل التكرار المعدل للفترات نكتب في Numeric Expression التعبير f/w

٤- نضغط على Ok فنعود للملف وقد قام بإضافة متغير mf يحتوي على التكرار المعدل للفترات

*ex3_19.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : L 10.0 Visible: 5 of 5 Variables

	L	U	f	w	mf
1	10.00	15.00	15.00	5.00	3.00
2	15.00	18.00	24.00	3.00	8.00
3	18.00	20.00	22.00	2.00	11.00
4	20.00	22.00	20.00	2.00	10.00
5	22.00	25.00	12.00	3.00	4.00
6	25.00	30.00	5.00	5.00	1.00

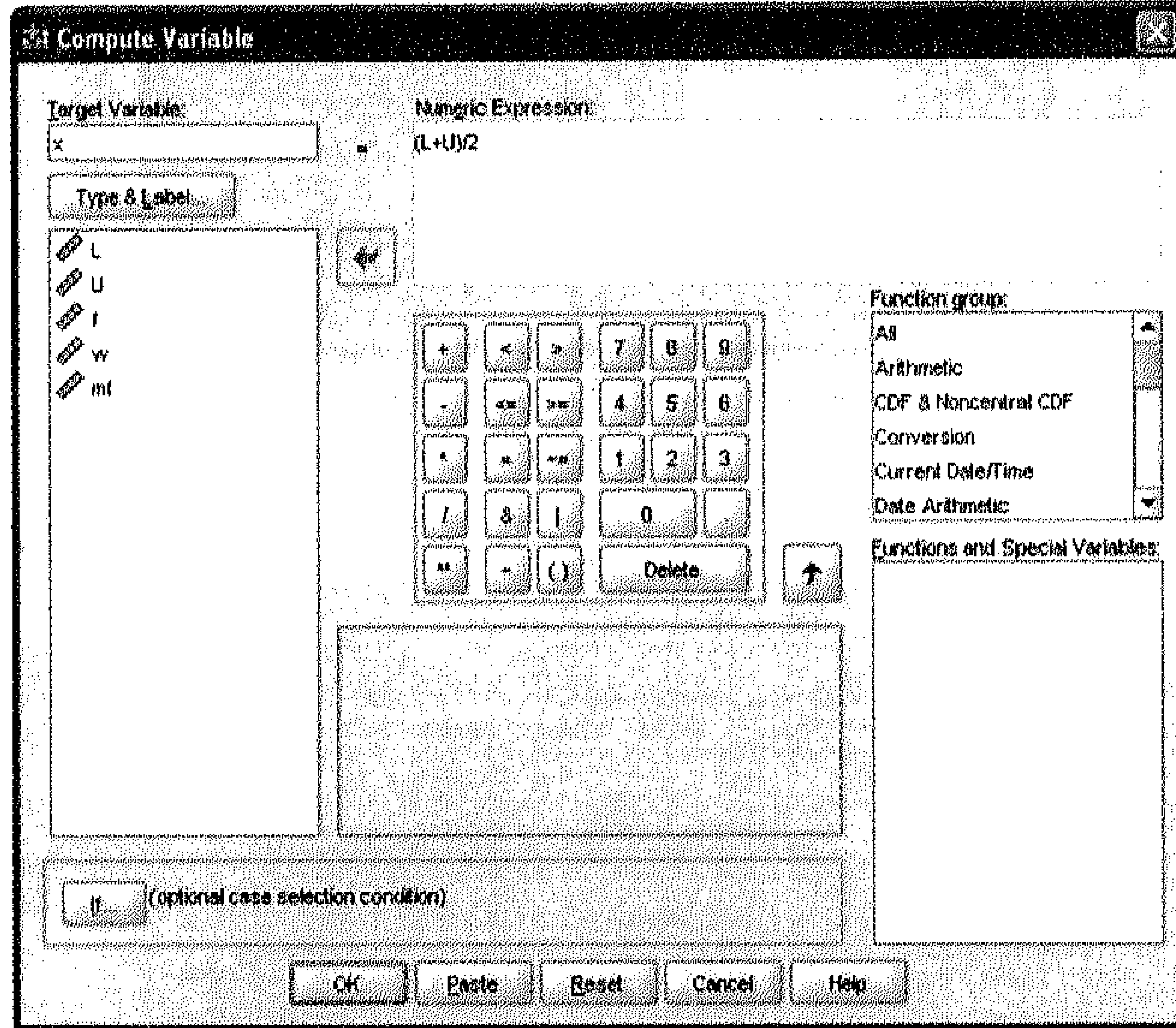
Data View Variable View

SPSS Statistics Processor is ready Weight On

رابعاً: تعيين مركز الفترات باتباع الخطوات التالية:

١- من قائمة Transform نختار Compute Variable

٢- تظهر شاشة جديدة بعنوان Compute Variable نكتب اسم المتغير الجديد *X* ويمثل مركز الفترات (L+U)/2 ونكتب تلك العلاقة في مربع Numeric Expression



٣- نضغط على Ok فنعود للملف وقد قام بإضافة متغير *X* يحتوي مركز الفترات

*ex3_19.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

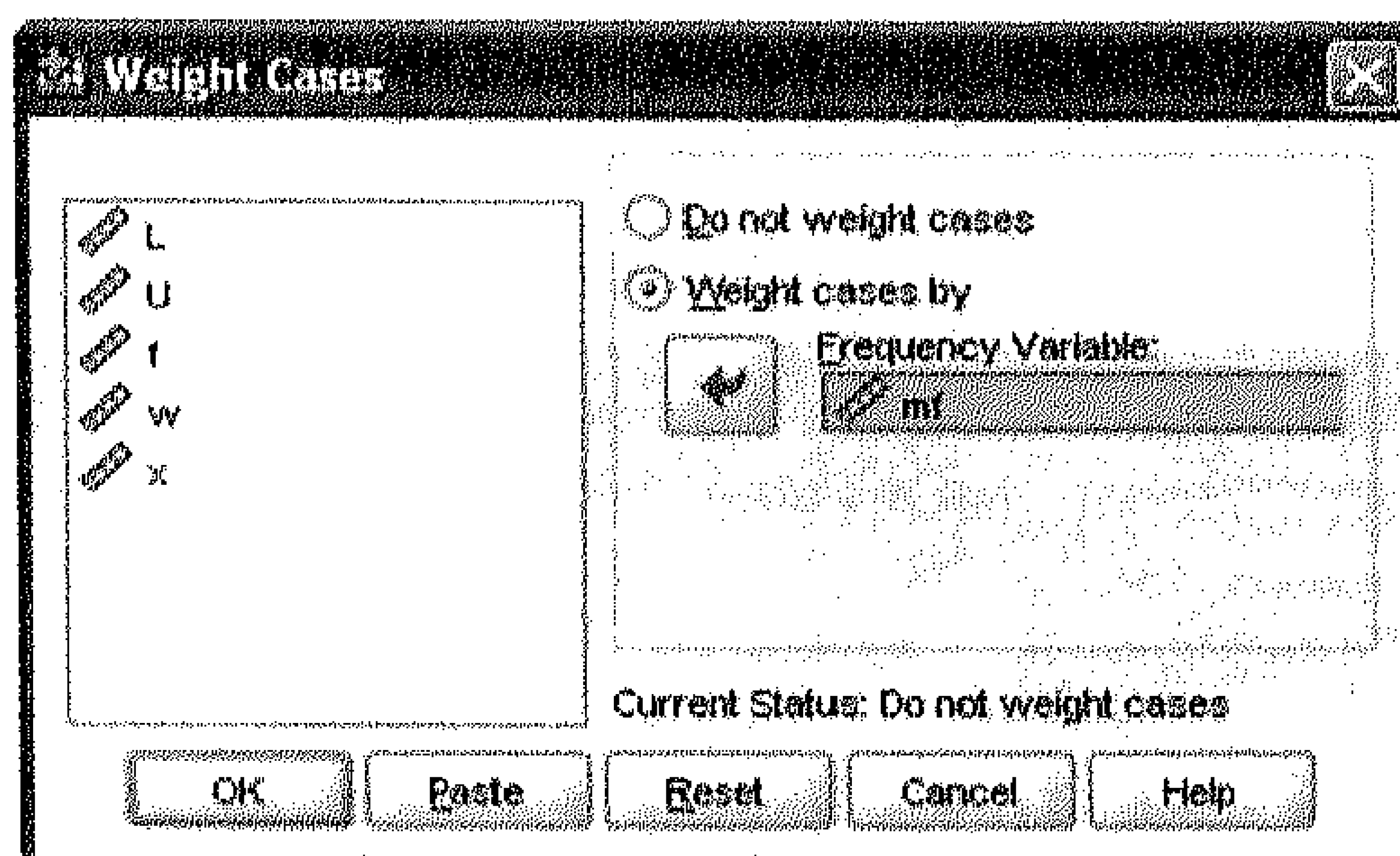
1 : L 10.0 Visible: 6 of 6 Variables

	L	U	f	w	mf	x	var
1	10.00	15.00	15.00	5.00	3.00	12.50	
2	15.00	18.00	24.00	3.00	8.00	16.50	
3	18.00	20.00	22.00	2.00	11.00	19.00	
4	20.00	22.00	20.00	2.00	10.00	21.00	
5	22.00	25.00	12.00	3.00	4.00	23.50	
6	25.00	30.00	5.00	5.00	1.00	27.50	
7							
8							

Data View Variable View

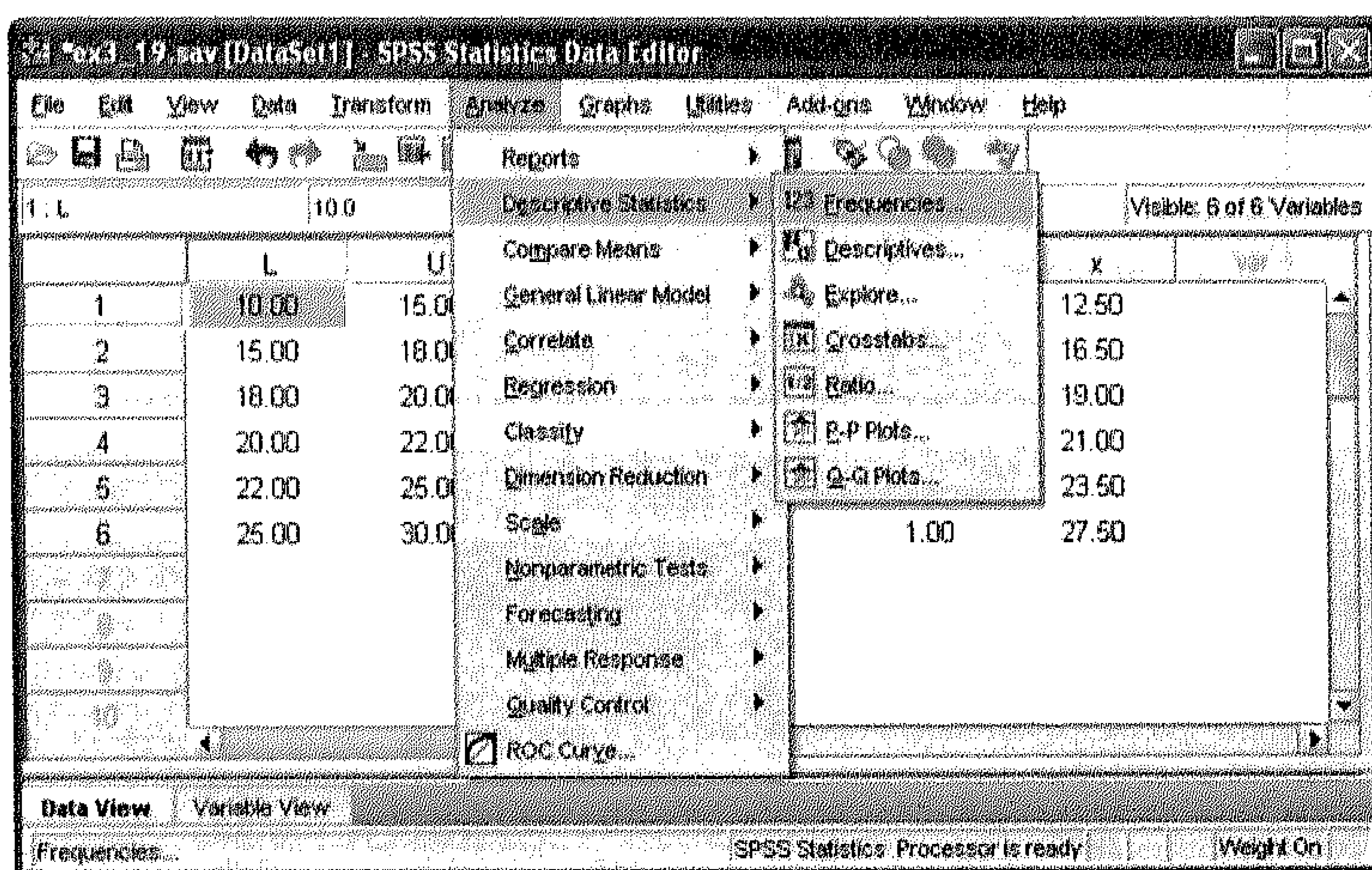
SPSS Statistics Processor is ready Weight On

خامساً: تعريف mf على أنها تكرار لمركز الفترات X وذلك باستخدام الأمر weight cases من قائمة Data



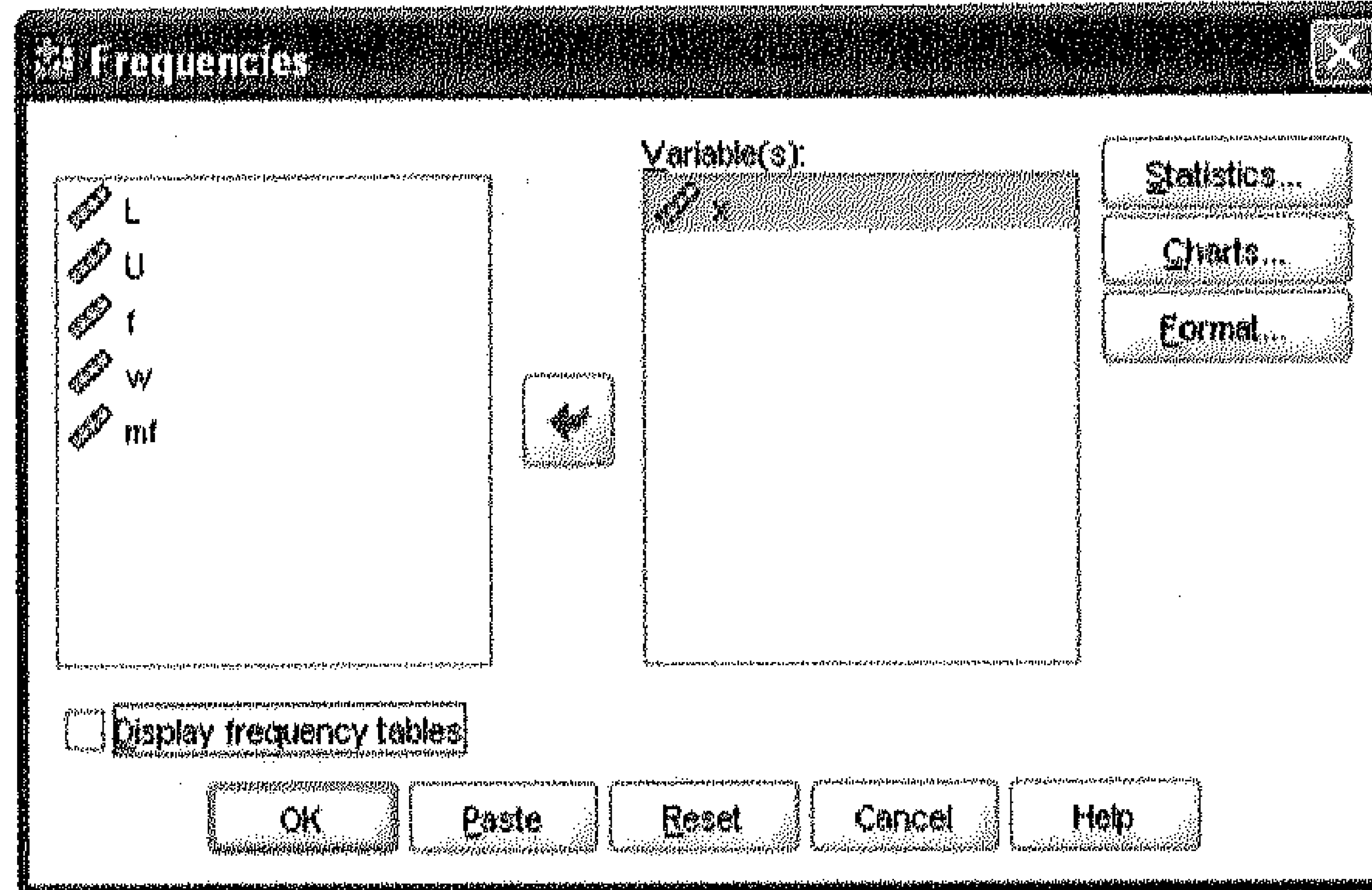
سادساً: تعيين المقاييس الإحصائية المطلوبة باتباع الخطوات التالية:

١- من قائمة Analyze نختار Descriptive Statistics ومن القائمة المنسدلة نختار Frequencies

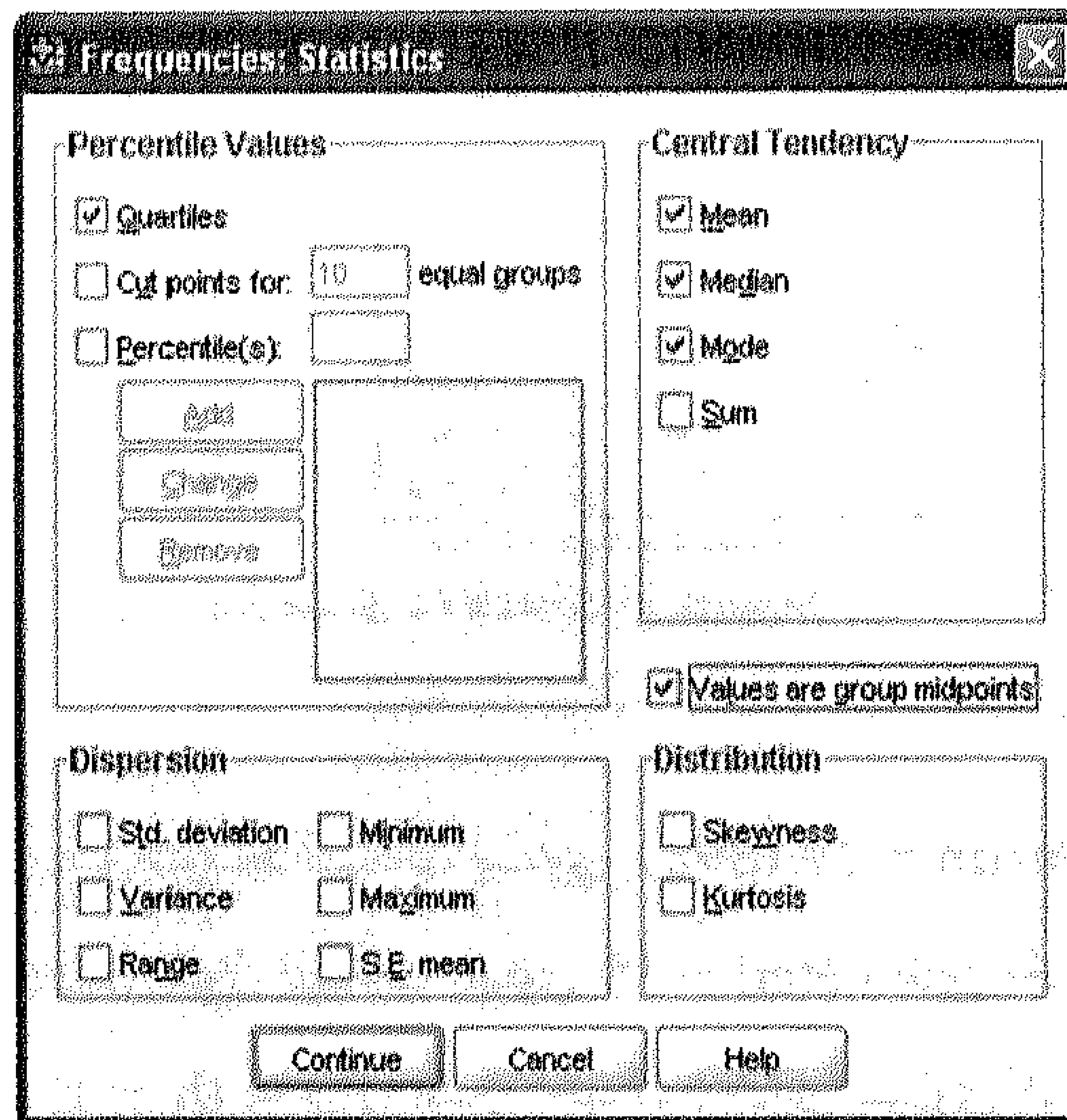


٢- تظهر نافذة بعنوان Frequencies

٣- من النافذة الجديدة ننقل المتغير X لقائمة variable(s)



٤- نضغط على الاختيار Statistics فتظهر شاشة جديدة



٥- نختار كلا من Mean, Median, Mode من قائمة Central Tendency ونختار Quartiles من قائمة Percentile values ويجب تحديد الاختيار Values are group midpoints لأن X يمثل مركز الفترات

٦- نضغط على Continue فنعود للشاشة السابقة فنختار Ok فتظهر النتائج التالية:

x

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	12.50	3	8.1	8.1	8.1
	16.50	8	21.6	21.6	29.7
	19.00	11	29.7	29.7	59.5
	21.00	10	27.0	27.0	86.5
	23.50	4	10.8	10.8	97.3
	27.50	1	2.7	2.7	100.0
	Total	37	100.0	100.0	

من الجدول السابق نجد أن العمود الأول يمثل مراكز الفترات والعمود الثاني هو التكرار المعدل والعمود الأخير هو التكرار النسبي التراكمي للتكرار المعدل.

Statistics

x

N	Valid	37
	Missing	0
Mean		19.1892
Median		19.3810[a]
Mode		19.00
Percentiles	25	17.0921[b]
	50	19.3810
	75	21.2679

a. Calculated from grouped data.

b. Percentiles are calculated from grouped data.

ونجد من الجدول السابق أن

عدد البيانات يساوى $n = 37$ ، المتوسط للبيانات يساوي $\bar{x} = Mean = 19.1892$، الوسيط للبيانات هو $Med = Median = 19.381$، المنوال للبيانات يساوي $Mod = 19$ ، المئين الخامس والعشرون الذي يساوي الربيع الأول يساوي $P_{25} = Q_1 = 17.0921$، المئين الخمسون الذي يساوي كلا من الربيع الثاني والوسيط يساوي $P_{50} = Q_2 = 19.381$، المئين الخامس والسبعون والذي يساوي الربيع الثالث يساوي $P_{75} = Q_3 = 21.2679$

أسئلة وتمارين (٣)

١- اشرح ما المقصود بمقاييس النزعة المركزية.

٢- ما الصفات التي يجب أن تتوفر في المقياس الجيد للنزعة المركزية؟

٣- عرف الوسيط والمنوال.

٤- ماذا نعني بأن المنحنى متماثل؟

٥- ما الفرق بين الوسط الحساب والوسيط؟

٦- عرف كلاً من الربيعات والعشيرات والمئينات.

٧- أي من مقاييس النزعة المركزية يمكن استخدامه مع البيانات الوصفية؟

٨- أوجد الوسط الحسابي للأعداد $1, 2, 3, \ldots, n$

٩- إذا كان الوسط الحسابي لدرجات 30 طالباً يساوي 40 فإذا أراد مدرس المادة زيادة الدرجات بمقدار 2 لكل قيمة فإن الوسط الحسابي للدرجات بعد التعديل يساوى.............

١٠- إذا كان الوسط الحسابي لدرجات 20 طالباً هي 40 والوسط الحسابي لدرجات 30 طالباً آخرين في نفس المقرر هو 45 فإن الوسط الحسابي لدرجات الخمسين طالباً بعد دمجهما معا هو...........

١١- واحد من المقاييس التالية ليس مقياساً للنزعة المركزية. ما هو؟

(أ) المتوسط الحسابي

(ب) المنوال

(ج) المئين

(د) الانحراف المتوسط

١٢- إذا كان الوسط الحسابي لعلامات 30 طالباً في مقرر الإحصاء يساوى 60 وعدلت العلامات وفق المعادلة التالية $y = 80 - 0.25\,x$ حيث x تمثل العلامات قبل التعديل، y العلامات بعد التعديل فإن الوسط الحسابي بعد التعديل هو ....

١٣- مجموع انحرافات القيم عن وسطها الحسابي يساوي ....

١٤- اكتب المصطلح المناظر للعبارات التالية:

- القيمة التي يقابلها أكبر تكرار
- القيمة التي تتوسط البيانات
- يعتبر من أفضل مقاييس النزعة المركزية لكن يعيبه أنه يتأثر بالقيم المتطرفة
- القيمة التي يوجد قبلها ربع البيانات وبعدها ثلاثة أرباع البيانات
- القيمة التي توجد قبلها واحد على عشرة من البيانات وبعدها باقي البيانات
- المقياس الذي لا يتأثر بإضافة قيمة ثابتة لجميع البيانات

١٥- أي من العبارات التالية صحيح وأيها خطأ

- الوسط الحسابي هو حالة خاصة من الوسط المرجح
- من عيوب المنوال أنه يتأثر بالقيم المتطرفة
- لا يمكن تعيين الوسيط للبيانات الوصفية
- الوسيط للبيانات 5 ,4 ,3 ,3 ,2 ,1 هو القيمة 3
- الوسط الحسابي للقيم $x_1, x_2, \dots, x_n$ يساوي الوسط التوافقي للقيم $\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_n}$
- البيانات التالية ليس لها منوال 1 ,1 ,1 ,1 ,1 ,1

١٦- البيانات التالية هي درجات أعمال السنة للطلاب في احد المقررات الدراسية والتي درجاتها من 10 درجات

الدرجة	4	5	6	7	8	9	10
عدد الطلاب	2	3	6	12	10	4	3

(أ) احسب متوسط الدرجات للطلاب

(ب) الوسيط لدرجات الطلاب

(ج) عين قيمة الوسط الهندسي والتوافقي للبيانات

١٧- فيما يلي التوزيع التكراري لأعمار 100مصباح كهربائي (بالساعة)

فئات العمر	0–	250–	500–	750–	1000–	1250–	1500–	1750–
عدد المصابيح	1	3	7	12	25	39	11	2

(أ) احسب الوسط الحسابي

(ب)احسب الوسيط والمنوال

(ج) أوجد الربيع الثالث والعشير السابع

(د) ادرس تماثل منحنى البيانات

١٨- فيما يلي التوزيع التكراري لأعمار مجموعة من البطاريات التي تستخدم في تشغيل الآلات الحاسبة الصغيرة (بالساعة)

فئات العمر	14–	614–	1214–	1814–	2414–	3014–
عدد البطاريات	23	7	9	2	4	5

(أ) احسب المنوال لأعمار البطاريات

(ب)احسب المئين الثالث والستين

(ج) احسب الوسط الهندسي لأعمار البطاريات

(د) احسب الوسط التوافقي للبيانات

١٩- فيما يلي التوزيع التكراري لدرجات الحرارة المئوية خلال 60 يوماً متتالية

درجات الحرارة	16–	18–	20–	22–	24–	26–	28–
عدد الأيام	4	7	10	18	9	7	5

(أ) احسب الوسط الحسابي والوسيط والمنوال

(ب)احسب العشير الثالث والمئين الخمسين

(ج) احسب الوسط التوافقي لدرجات الحرارة

(د) ادرس تماثل منحنى درجات الحرارة

٢٠- البيانات التالية تمثل الأجور اليومية لخمسة وعشرون عاملاً

الأجور اليومية	3	3.5	4	5	5.5	6	المجموع
عدد العمال	5	2	4	3	6	5	25

أوجد

(أ) متوسط الأجور اليومية

(ب) الوسيط للأجور اليومية

(ج) المنوال للأجور اليومية

٢١- البيانات التالية تمثل أوزان 10 طلاب 40 ,64 ,60 ,52 ,56 ,41 ,59 ,48 ,65 ,45 والمطلوب إيجاد الوسط والوسيط لهذه البيانات.

٢٢- في البرنامج الموحد للكليات الصحية كانت درجات الطلاب في امتحان 1060 إحص من عشرة درجات كالتالي:

الدرجة	4	5	6	7	8	9	10
عدد الطلاب	2	8	13	35	21	16	5

أوجد المنوال والوسط الحسابي لهذه البيانات.

٢٣- البيانات التالية تمثل أعمار عينة من المقيمين بأحد الفنادق بمكة المكرمة في أحد مواسم الحج

الفترات	20 – 24	25 – 29	30 – 34	35 – 39	40 – 44	المجموع
التكرار	7	13	21	6	3	50

(أ) أوجد

- المتوسط لعمر المقيمين بالفندق
- الربيع الثالث والمئين الخمسين
- المنوال لبيانات العمر

(ب) ادرس تماثل توزيع العمر

٢٤- إذا كان لدينا البيانات التالية

الفترات	30 –	40 –	50 –	60 –	70 –	المجموع
التكرار	2	9	21	11	7	50

أوجد

(أ) الوسط التوافقي للبيانات

(ب) المنوال للبيانات

(ج) المئين الثالث والأربعين

٢٥- إذا كان لدينا ثلاثة عينات أحجامها على التوالي هي $n_1 = 15, n_2 = 20, n_3 = 25$ وكانت أوساطها الحسابية $\bar{x}_1 = 45, \bar{x}_2 = 75, \bar{x}_3 = 60$ فأوجد الوسط الحسابي للعينة الناتجة من دمج تلك العينات معاً.

٢٦- البيانات التالية تمثل الأجور الأسبوعية لمائة عامل موزعة كالتالي

الأجور	60 – 69	70 – 79	80 – 89	90 – 99	100 – 109	المجموع
عدد العمال	6	12	47	25	10	100

(أ) أوجد

- الوسط الهندسي لأجور العمال
- الوسيط والمنوال بيانياً

(ج) ادرس تماثل توزيع الأجور الأسبوعية للعمال

٢٧- البيانات التالية تمثل أطوال 40 طالباً في احدي الشعب الدراسية لمقرر 1080 إحص بكلية العلوم موزعه كما يلي

فترات الطول	157 – 159	160 – 162	163 – 165	166 – 168	169 – 171
عدد الطلاب	5	7	12	9	7

أوجد

(أ) متوسط طول الطلاب

(ب) الوسيط بيانياً

(ج) المنوال لأطوال الطلاب

٢٨- البيانات التالية تمثل الأجور الشهرية لمنسوبي إحدى المؤسسات الخاصة موزعة كالتالي

الأجر الشهري	1500 –	2000 –	2500 –	3000 –	3500 –	4000 –	4500 –
عدد الأفراد	8	3	25	17	10	4	3

والمطلوب تعيين

(أ) الربيع الثاني والربيع الثالث

(ب) العشير الخامس والعشير السابع

(ج) المئين الثالث والثلاثين، المئين الخمسين بيانياً

(د) قارن بين كل من الوسط الحسابي، الوسط التوافقي، الوسط الهندسي

٢٩- فيما يلي التوزيع التكراري لأعمار 100 جهاز حاسب آلي بمعامل كلية العلوم والدراسات الإنسانية (باليوم)

فئات العمر	0–	250–	500–	750–	1000–	1250–
عدد الأجهزة	7	6	11	12	25	39

(أ) احسب الوسط الحسابي

(ب) احسب الوسيط والمنوال

(ج) أوجد الربيع الثالث والعشير السابع

(د) ادرس تماثل منحنى البيانات

٣٠- فيما يلي التوزيع التكراري لدرجات مجموعة من الطلاب في مقرر 1080 كيم

فئات الدرجات	40–	50–	60–	70–	80–	90–
عدد الطلاب	3	7	14	12	9	5

احسب

(أ) المنوال لدرجات الطلاب

(ب) المئين الثالث والستين

(ج) الوسط الهندسي لدرجات الطلاب

(د) الوسط التوافقي للبيانات

الفصل الرابع

مقاييس التشتت

MEASURES OF DISPERSION

٤–١ مقدمة

لاحظنا أن مقاييس النزعة المركزية لمجموعة من البيانات تعطى فكرة عامة عن هذه البيانات، فهي تبين مدى تمركز البيانات حول قيمة معينة. وفى بعض الأحيان تكون البيانات أو المشاهدات قريبة من هذه القيمة وفى أحيان أخرى تكون منتشرة في مدى أوسع حولها، ولتوضيح ذلك نفرض أننا اخترنا مجموعتين من الطلاب كل منها تحتوي على عشرة طلاب وقمنا بسؤال كل منهم عن الدرجة التي حصل عليها في مادة معينة فكانت درجاتهم كما يلي:

المجموعة الثانية	المجموعة الأولى
100 93 85 77 67 62 58 54 77 42	77 77 77 74 73 71 68 67 66 65

وبحساب كل من الوسط الحسابي والوسيط والمنوال لكل مجموعة نجدها:

المجموعة الثانية	المجموعة الأولى
$\bar{x} = 71.5$ Median = 72 Mode = 77	$\bar{x} = 71.5$ Median = 72 Mode = 77

وبالنظر إلى هذه المقاييس فإننا نلاحظ لأول وهلة تساوي مقاييس النزعة المركزية لهاتين المجموعتين. ولكن بنظرة أكثر عمقا نجد أن هناك فرقاً بين المجموعتين وهو أن درجات المجموعة الثانية تنتشر في مدى أوسع من المدى الذي تنتشر فيه درجات المجموعة الأولى فالاختلاف أو التغير (variability) بين المشاهدات صفة لا تظهرها مقاييس النزعة المركزية ويسمى هذا الاختلاف أو التغير بالتشتت (dispersion). وتسمى المقاييس المستخدمة في قياس التشتت بمقاييس التشتت (measures of dispersion) ومن بين مقاييس التشتت المدى، نصف المدى الربيعي والانحراف المعياري.

٤-٢ المدى (Range)

يعتبر المدى أبسط مقاييس التشتت وهو الفرق بين أكبر مشاهدة وأقل مشاهدة ويرمز له بالرمز R

$$R = X_{max} - X_{min}$$

ونجد في حالة البيانات المفردة أن X_{max} تمثل أكبر قيمة في البيانات، X_{min} تمثل أقل قيمة في البيانات. لكن في حالة البيانات المبوبة للمتغير الكمي المتصل فإن X_{max} تمثل مركز آخر فترة في الجدول، X_{min} تمثل مركز أول فترة في الجدول التكراري.

فمثلا بالنسبة لمثال درجات الطلاب السابق نجد أن

$R_1 = 77 - 65 = 12$ درجة

$R_2 = 100 - 42 = 58$ درجة

نلاحظ أن $R_2 > R_1$ وهذا يعنى أن مشاهدات المجموعة الثانية أكثر تشتتاً من مشاهدات المجموعة الأولى وهذا يعنى أن مشاهدات المجموعة الأولى أكثر تجانساً من مشاهدات المجموعة الثانية.

مثال (٤-١)

فيما يلي التوزيع التكراري لإنتاج أحد المحاصيل (بآلاف الكيلوجرامات) من مجموعة من قطع الأراضي الزراعية المتساوية:

فئات الإنتاج	26–	31–	36–	41–	46–	51–	56–	61–
عدد القطع	4	5	23	58	61	30	4	3

المطلوب إيجاد المدى لكمية الإنتاج.

الحل

لتعيين مدى البيانات يجب أولا تعيين مركز الفترات

الفترات	التكرار f_i	مركز الفترات x_i
26 – 31	4	28.5
31 – 36	5	33.5
36 – 41	23	38.5
41 – 46	58	43.5
46 – 51	61	48.5
51 – 56	30	53.5
56 – 61	4	58.5
61 – 66	3	63.5

من الجدول السابق نجد أن

$$X_{min} = 28.5, \ X_{max} = 63.5$$

ومنها فإن المدى لكمية الإنتاج لقطع الأراضي هو

$$R = X_{max} - X_{min} = 63.5 - 28.5 = 35$$

كما نرى فإن المدى يمتاز بأنه سهل الحساب، ولكنه يتأثر كثيراً بالقيم الشاذة وذلك بسبب اعتماده على مشاهدتين اثنتين فقط هما أكبر وأصغر قيمة، كما أنه لا يمكن تعيينه للبيانات الوصفية.

٤-٣ نصف المدى الربيعي (Semi- inter quartile Range)

المدى الربيعي هو الفرق بين الربيع الثالث Q_3 والربيع الأول Q_1. فإن نصف المدى الربيعي ويرمز له بالرمز Q ويمكن حسابه من العلاقة الآتية:

$$Q = \frac{Q_3 - Q_1}{2}$$

مثال (٤-٢)

تم اختيار عينة من 72 مسماراً منتجة بواسطة ماكينة تعمل آليا فكانت أطوالها (لأقرب 0.01 سنتيمتر) هي

فئات الطول	4.4–	4.5–	4.6–	4.7–	4.8–	4.9–
عدد المسامير	5	8	14	20	17	8

احسب نصف المدى الربيعي.

الحل

لتعيين نصف المدى الربيعي يجب أولا تعيين كل من الربيع الأول والربيع الثالث. لذا سوف نقوم بتكوين جدول التوزيع التكراري المتجمع الصاعد

التكرار المتجمع الصاعد	الحدود العليا للفئات
0	أقل من4.4
5	أقل من4.5
13	أقل من4.6
27	أقل من 4.7
47	أقل من4.8
64	أقل من 4.9
72	أقل من5.0

الربيع الأول

موقع الربيع الأول

$$R_1 = \frac{1 \times n}{4} = \frac{72}{4} = 18$$

ومن جدول التوزيع التكراري المتجمع الصاعد نجد أن الفئة التي تحتوى الربيع الأول هي4.7 – 4.6، ومنها فإن

$$L = 4.6, \qquad h = 0.1, \qquad F_1 = 13, \quad F_2 = 27$$

فإن الربيع الأول هو

$$Q_1 = L + \left(\frac{R_1 - F_1}{F_2 - F_1}\right) \times h$$

$$= 4.6 + \left(\frac{18 - 13}{27 - 13}\right) \times 0.1$$

$$= 4.6357$$

الربيع الثالث

موقع الربيع الثالث

$$R_3 = \frac{3 \times n}{4} = \frac{3 \times 72}{4} = 54$$

ومن جدول التوزيع التكراري المتجمع الصاعد نجد أن الفئة التي تحتوى الربيع الثالث هي$4.9 - 4.8$، ومنها فإن:

$$L = 4.8,\ \ h = 0.1,\ \ F_1 = 47,\ \ F_2 = 64$$

فإن قيمة الربيع الثالث هي

$$Q_3 = L + \left(\frac{R_3 - F_1}{F_2 - F_1}\right) \times h$$

$$= 4.8 + \left(\frac{54 - 47}{64 - 47}\right) \times 0.1$$

$$= 4.841$$

مما سبق فإن نصف المدى الربيعي هو

$$Q = \frac{Q_3 - Q_1}{2} = \frac{4.841 - 4.6357}{2} = 0.1027$$

إن استخدام نصف المدى الربيعي يخلصنا من العيوب التي توجد في المدى، حيث إنه لا يتأثر بالقيم الشاذة كما أنه يمكن إيجاده في حاله التوزيعات التكرارية للمتغيرات المتصلة.

٤-٤ التباين والانحراف المعياري (Variance and Standard Deviation)

يعتبر التباين من أهم مقاييس التشتت وأكثرها انتشاراً، وهو يختلف عن كل من المدى ونصف المدى الربيعي في أن كلاً من المقياسين الأخيرين يعتمد على قيمتين اثنتين فقط بينما يعتمد التباين على كل المشاهدات وبالتحديد على انحرافات المشاهدات عن وسطها الحسابي، ويمكن حساب التباين كما يلي:

(أ) البيانات المفردة (Raw Data)

بفرض أن لدينا عينة من المشاهدات $x_1, x_2, \dots, x_n$ وسطها الحسابي هو $\bar{x}$ فإن انحرافات المشاهدات عن وسطها الحسابي هو $x_1 - \bar{x},\ x_2 - \bar{x},\ \ x_3 - \bar{x}, \dots,\ \ \ x_n - \bar{x}$ ومجموع هذه الانحرافات يساوى صفراً.

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

ومربعــات هــذه الانحرافــات هــي $(x_1-\bar{x})^2, (x_2-\bar{x})^2, \ (x_3-\bar{x})^2, \dots, (x_n-\bar{x})^2$ ومجمــوع مربعــات انحرافات المشاهدات عن وسطها دائما يكون مقدار موجب إلا إذا كانت جميع المشاهدات متماثلة في هذه الحالة فإنه يساوى صفراً.

$$\sum_{i=1}^{n}(x_i-\bar{x})^2 \geq 0$$

ومتوسط مربعات انحرافات المشاهدات عن الوسط يسمى التباين (variance) ويرمز له بالرمز s^2 أي أن:

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

والجذر التربيع للتباين يسمى الانحراف المعياري ويرمز له بالرمز s أي أن

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

مثال (٤-٣)

اختيرت عينة من 10 مكالمات دولية بعد الساعة 11:00 مساءً وقبل الساعة 8:00 صباحاً فكانت أطوالها بالدقائق هي:,10, 20, 6, 12, 15, 8, 4, 9, 3,13 احسب الانحراف المعياري والتباين لأطوال المكالمات.

الحل

لحساب التباين والانحراف المعياري يجب أولاً تعيين الوسط الحسابي

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{10+20+6+\cdots+3+13}{10}=\frac{100}{10}=10$$

ويمكن حساب التباين من العلاقة:

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

وبتكوين الجدول التالي:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
10	0	0
20	10	100
6	-4	16
12	2	4
15	5	25
8	-2	4
4	-6	36
9	-1	1
3	-7	49
13	3	9
المجموع	0	244

$$\Longrightarrow s^2 = \frac{244}{10-1} = \frac{244}{9} = 27.111$$

وبالتالي فإن الانحراف المعياري يساوي

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$= \sqrt{27.111} = 5.207$$

ويمكن إجراء بعض العمليات الجبرية ووضع الصيغة السابقة للتباين في صورة جديدة تجعل العمليات الحسابية أسهل كثيراً وهذه الصيغة هي:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \bar{x}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\,\bar{x}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2\right]$$

ويمكن حساب التباين بهذه العلاقة للمثال السابق

x_i	x_i^2
10	100
20	400
6	36
12	144
15	225
8	64
4	16
9	81
3	9
13	169
$\sum x_i = 100$	$\sum x_i^2 = 1244$

فإن التباين يعطى كالتالي:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2\right]$$

$$= \frac{1}{10-1}\left[1244 - 10\left(\frac{100}{10}\right)^2\right]$$

$$= \frac{1}{9}[\,1244 - 1000] = \frac{244}{9} = 27.111$$

وهى نفس القيمة السابقة وكذلك يكون الانحراف المعياري:

$$s = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2\right]}$$

$$= \sqrt{27.111} = 5.207$$

(ب) البيانات المبوبة (Grouped Data)

في هذه الحالة سوف يتم حساب التباين من العلاقة:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i-\bar{x})^2 f_i$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2 f_i - n\left(\frac{\sum_{i=1}^{k} x_i f_i}{n}\right)^2\right]$$

حيث x_i هى القيم الممكنة للمتغير في حالة التوزيعات التكرارية للمتغيرات المتقطعة أو هي مراكز الفترات في حالة التوزيعات التكرارية للمتغيرات المتصلة. f_i هو التكرار المقابل للقيمة x_i وكذلك n هى مجموع التكرارات، k عدد القيم(الفترات).

مثال (٤-٤)

لعينة من 500 أسرة اختيرت من إحدى المدن توفرت لدينا البيانات التالية عن عدد أفراد الأسرة

عدد أفراد الأسرة	2	3	4	5	6	7	8	9
عدد الأسر	98	118	101	95	50	20	10	8

أحسب التباين والانحراف المعياري لعدد أفراد الأسرة.

الحل

المتغير هو عدد أفراد الأسرة وهو متغير كمي متقطع وبالتالي سوف نحسب التباين

باستخدام القانون الأول:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i-\bar{x})^2 f_i$$

أولاً يجب تعيين المتوسط للبيانات

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i = \frac{2021}{500} = 4.042$$

ثانياً بتعيين الجدول التالي:

x_i	f_i	$x_i f_i$	$(x_i-\bar{x})$	$(x_i-\bar{x})^2 f_i$
2	98	196	-2.042	408.6369
3	118	354	-1.042	128.1202
4	101	404	-0.042	0.178164
5	95	475	0.958	87.18758
6	50	300	1.958	191.6882
7	20	140	2.958	174.9953
8	10	80	3.958	156.6576
9	8	72	4.958	196.6541
المجموع	500	2021		1344.118

من الجدول السابق نجد أن

$$s^2=\frac{1}{n-1}\sum_{i=1}^{k}(x_i-\bar{x})^2 f_i$$

$$=\frac{1344.118}{500-1}=2.694$$

باستخدام القانون الثاني

$$s^2=\frac{1}{n-1}\left[\sum_{i=1}^{k}x_i^2 f_i-n\left(\frac{\sum_{i=1}^{k}x_i f_i}{n}\right)^2\right]$$

x_i	f_i	$x_i f_i$	$x_i^2 f_i$
2	98	196	392
3	118	354	1062
4	101	404	1616
5	95	475	2375
6	50	300	1800
7	20	140	980
8	10	80	640
9	8	72	648
المجموع	500	2021	9513

من الجدول السابق نجد أن التباين يساوى

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2 f_i - n\left(\frac{\sum_{i=1}^{k} x_i f_i}{n}\right)^2\right]$$

$$= \frac{1}{500-1}\left[9513 - 500\left(\frac{2021}{500}\right)^2\right]$$

$$= 2.694$$

والانحراف المعياري

$$s = \sqrt{2.694} = 1.641$$

مثال (٤-٥)

الجدول التكراري الآتي يبين توزيع 100 عمود من الصلب حسب طولهم (إلى أقرب 0.01 بوصة)

فئات الطول	3.8–	3.9–	4.0–	4.1–	4.2–	4.3–	4.4–
التكرار	3	8	14	19	28	18	10

احسب التباين والانحراف المعياري لطول العمود

الحل:

المتغير هو طول عمود الصلب وهو متغير كمي متصل. وسوف نقوم بحساب التباين لطول العمود من العلاقة التالية:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2 f_i - n\left(\frac{\sum_{i=1}^{k} x_i f_i}{n}\right)^2\right]$$

فئات الطول	التكرار f_i	مركز الفترات i	$x_i f_i$	$x_i^2 f_i$
3.8 – 3.9	3	3.85	11.55	44.4675
3.9 – 4.0	8	3.95	31.6	124.82
4.0 – 4.1	14	4.05	56.7	229.635
4.1 – 4.2	19	4.15	78.85	327.2275
4.2 – 4.3	28	4.25	119	505.75
4.3 – 4.4	18	4.35	78.3	340.605
4.4 – 4.5	10	4.45	44.5	198.025
المجموع	100		420.5	1770.53

من الجدول السابق فإن:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{k} x_i^2 f_i - n\left(\frac{\sum_{i=1}^{k} x_i f_i}{n}\right)^2\right]$$

$$= \frac{1}{100-1}\left[1770.53 - 100\left(\frac{420.5}{100}\right)^2\right]$$

$$= 0.0235$$

والانحراف المعياري

$$s = \sqrt{0.0235} = 0.1533$$

بعض خصائص التباين:

١- إذا كانت $x_1, x_2, \ldots x_n$ مجموعة من القيم وسطها الحسابى $\bar{x}$ وتباينها s^2 وكانت a, b مقدارين ثابتين فإن

(أ) تباين البيانات $ax_1,\ ax_2, \ldots, ax_n$ هو $a^2 s^2$

(ب) تباين البيانات $ax_1 \pm b,\ ax_2 \pm b, \ldots, ax_n \pm b$ هو $a^2 s^2$

٢- وإذا كان لدينا مجموعتان عدد عناصرهما يساوي n_1, n_2 والمتوسط الحسابي لهما يساوي $\bar{x}_1,\ \bar{x}_2$ وتباينهما $s_1^2,\ s_2^2$ على التوالي فإن التباين للمجموعة الكلية الناتجة من دمج تلك المجموعات معاً يعطى بالعلاقة التالية:

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

مثال(٤-٦)

إذا كان متوسط أجور 36 عامل بإحدى الشركات هو 1800 ريال سعودي بتباين مقداره 400 ريال ومتوسط أجور 25 فني بنفس الشركة هو 2500 ريال بتباين مقداره 1600 ريال فما هو الوسط الحسابي والتباين للمجموعة الكلية بعد دمج الفئتين من منسوبي الشركة معاً.

الحل

من البيانات المعطاة فإن

أجور الفنيين	أجور العمال
$n_2 = 25$ $\bar{x}_2 = 2500$ $s_2^2 = 1600$	$n_1 = 36$ $\bar{x}_1 = 1800$ $s_1^2 = 400$

المتوسط للمجموعة الكلية

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$= \frac{36 \times 1800 + 25 \times 2500}{25 + 36}$$

$$= 2086.885$$

التباين للمجموعة الكلية

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(36 - 1) \times 400 + (25 - 1) \times 1600}{25 + 36 - 2}$$

$$= 888.1356$$

٤–٥ الوحدات القياسية (Standard Units)

من المفيد أحيانا أن نقارن بين مشاهدتين تم اختيارهما من مجموعتين مختلفتين من حيث الوسط الحسابي والانحراف المعياري. ويستخدم في ذلك ما يسمى بالوحدات القياسية وقد تسمى بالدرجات المعيارية، ويرمز لها بالرمز z . فإذا كان لدينا مجموعة من البيانات (المشاهدات) $x_1, x_2, x_3, \dots, x_n$ متوسطها الحسابى $\bar{x}$ وانحرافها المعياري هو s فإن الدرجة المعيارية للقيمة x_i تحسب من العلاقة التالية:

$$z_i = \frac{x_i - \bar{x}}{s}, \quad i = 1, 2, \dots, n$$

مثال(٤–٧)

إذا كانت درجة طالب في أحد الاختبارات هي 65 ودرجته في اختبار ثاني هي 29 ومن درجات طلاب المجموعتين التي ينتمي إليها الطالب تبين أن

الاختبار الثاني	الاختبار الأول
$\bar{x} = 20$ s = 5	$\bar{x} = 50$ s =10

ففي أي من المادتين كان أداء الطالب أفضل؟

الحل

لا يمكن مقارنة درجات الطالب هكذا لذا يجب تحويلها إلى درجات معيارية لكي نستطيع مقارنتها.

الدرجة المعيارية لدرجة الطالب في الاختبار الأول:

$$z_1 = \frac{x_1 - \bar{x}_1}{s_1} = \frac{65 - 50}{10} = 1.5$$

الدرجة المعيارية لدرجة الطالب في الاختبار الثاني:

$$z_2 = \frac{x_2 - \bar{x}_2}{s_2} = \frac{29 - 20}{5} = 1.8$$

ويمكن من ذلك الاستدلال على أن أداء الطالب في الاختبار الثاني أفضل من أدائه في الاختبار الأول.

٤-٦ التشتت النسبي (Relative Dispersion)

إذا أردنا أن نقارن بين تشتت مجموعتين أو أكثر من البيانات فإننا لا نستخدم مقاييس التشتت السابق ذكرها مباشرة وذلك بسبب:

(أ) أن بيانات المجموعات المختلفة قد تكون مقاسة بوحدات مختلفة كأن تكون مجموعة مقاسة بالسنتيمتر والأخرى مقاسة بالكيلو متر.

(ب) أن تكون الأوساط الحسابية للمجموعات مختلفة.

في هذه الحالات يجب استخدام ما يسمى بمقاييس التشتت النسبي وأهم هذه المقاييس هو معامل الاختلاف (coefficient of variation) ويرمز له بالرمز c.v. ومعامل الاختلاف عبارة عن نسبة مئوية ويحسب من العلاقة:

$$c.v. = \frac{s}{\bar{x}} \times 100$$

ومعامل الاختلاف عديم الوحدة أي ليس له وحدة قياس.

مثال (٤-٨)

فيما يلي بيانات عن أوزان (X) وأطوال (Y) مجموعة من 120 عاملاً تم اختيارهم من أحد المصانع:

فئات الوزن	50 –	55–	60 –	65 –	70 –	75 –	80 –
عدد العمال	8	12	20	30	25	15	10

فئات الطول	150 –	155–	160 –	165 –	170 –	175 –	180 –	185–
عدد العمال	7	10	20	25	18	17	13	10

قارن بين تشتت الأوزان وتشتت الأطوال باستخدام معامل الاختلاف.

الحل

بالنسبة للأوزان(X)

فترات الوزن	التكرار f_i	مركز الفترات x_i	$x_i f_i$	$x_i^2 f_i$
50 – 55	8	52.5	420	22050
55 – 60	12	57.5	690	39675
60 – 65	20	62.5	1250	78125
65 – 70	30	67.5	2025	136687.5
70 – 75	25	72.5	1812.5	131406.3
75 – 80	15	77.5	1162.5	90093.75
80 – 85	10	82.5	825	68062.5
المجموع	120		8185	566100.05

الوسط الحسابي:

$$\bar{x}_1 = \frac{1}{n_1}\sum_{i=1}^{k_1} x_i = \frac{8185}{120} = 68.208$$

الانحراف المعياري:

$$s_1^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{k_1} x_i^2 f_i - n_1\left(\frac{\sum_{i=1}^{k_1} x_i f_i}{n_1}\right)^2\right]$$

$$= \frac{1}{120 - 1}\left[566100.05 - 120\left(\frac{8185}{120}\right)^2\right] = 65.6709$$

$$\Longrightarrow s_1 = \sqrt{65.6709} = 8.10376$$

معامل الاختلاف:

$$c.v._1 = \frac{s_1}{\bar{x}_1} \times 100 = \frac{8.10376}{68.208} \times 100 = 11.8809\%$$

بالنسبة للأطوال(Y)

فترات الطول	f_i	y_i	$y_i f_i$	$y_i^2 f_i$
150 – 155	7	152.5	1067.5	162793.8
155 – 160	10	157.5	1575	248062.5
160 – 165	20	162.5	3250	528125
165 – 170	25	167.5	4187.5	701406.3
170 – 175	18	172.5	3105	535612.5
175 – 180	17	177.5	3017.5	535606.3
180 – 185	13	182.5	2372.5	432981.3
185 – 190	10	187.5	1875	351562.5
المجموع	120		20450	3496150.2

الوسط الحسابي:

$$\bar{y} = \frac{1}{n_2}\sum_{i=1}^{k_2} y_i = \frac{20450}{120} = 170.4167$$

الانحراف المعياري:

$$s_2^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{k_1} y_i^2 f_i - n_2\left(\frac{\sum_{i=1}^{k_2} y_i f_i}{n_2}\right)^2\right]$$

$$= \frac{1}{120 - 1}\left[3496150.2 - 120\left(\frac{20450}{120}\right)^2\right] = 93.5241$$

$$\Longrightarrow s_2 = \sqrt{93.5241} = 9.67079$$

معامل الاختلاف:

$$c.v._2 = \frac{s_2}{\bar{y}} \times 100 = \frac{9.67069}{170.4167} \times 100 = 5.6747\%$$

ومما سبق نجد أن معامل الاختلاف للأوزان أكبر من معامل الاختلاف للأطوال وبالتالي فإن التشتت النسبي للأوزان أكبر من التشتت النسبي للأطوال رغم أن الانحراف المعياري للأوزان أقل من الانحراف المعياري للأطوال.

٤-٧ متباينة تشيبيشف (Chebychev's Inequality)

سنتعرض لمتباينة شهيرة تملك قيمة تطبيقيه محدودة ، لكنها ذات قيمة نظرية كبيرة لأنها تستخدم أساساً لإثبات مجموعة من المبرهنات المهمة في الاحتمالات وتطبيقاتها، وهذه المتباينة تسمي بمتباينة تشيبيشف.

متباينة تشيبيشف:

إذا كان $x_1, x_2, \dots, x_n$ مجموعة من البيانات للمتغير العشوائي X بمتوسط $\bar{x}$ وتباين منتهي s^2، فان احتمال أن يأخذ المتغير العشوائي قيماً في الفترة $(\bar{x} - rs, \bar{x} + rs)$ لا يقل عن $1 - \frac{1}{r^2}$ حيث $r > 1$.

وهكذا نجد أن متباينة تشيبيشف يمكن تطبيقها على أي مجموعة من البيانات بشرط معرفة متوسطها الحسابي وانحرافها المعياري ولا نحتاج لمعرفة البيانات نفسها فباستخدام هذه المتباينة يمكن تعيين نسبة البيانات التي تقع في فترة طولها $2rs$ ومركزها متوسط البيانات $\bar{x}$.

مثال (٤-٩)

إذا كان لدينا مجموعة من البيانات متوسطها الحسابي 5 وتباينها هو 16 فأوجد الفترة التي يقع فيها ما لا يقل عن 60% من البيانات لهذه المجموعة.

الحل

من البيانات المعطاة نجد أن $\bar{x} = 5, s^2 = 16 , 1 - \frac{1}{r^2} = 0.6$

من متباينة تشيبيشف نجد أن الفترة التي يقع فيها ما لا يقل عن $1 - \frac{1}{r^2}$ هي $(\bar{x} - rs, \bar{x} + rs)$ وبالتالي فإن:

$$1 - \frac{1}{r^2} = 0.6 \Rightarrow \frac{1}{r^2} = 0.4 \Rightarrow r^2 = \frac{10}{4} = 2.5$$

$$\Rightarrow r = \sqrt{2.5} = 1.5811$$

وبالتالي فإن حدود الفترة هي:

$$\bar{x} - rs = 5 - 1.5811(4) = -1.3244,$$

$$\bar{x} + rs = 5 + 1.5811(4) = 11.3244$$

وبالتالي فإن ما لا يقل عن 60% من البيانات يقع في الفترة (11.3244 ,1.3244−).

مثال (٤-١٠)

إذا كان لدينا مجموعة من العمال بإحدى الشركات الخاصة متوسط أعمارهم 35 سنة والتباين لأعمارهم هو 4 سنوات فأوجد نسبة العمال بتلك الشركة التي تقع أعمارهم في الفترة (38 ,32) سنة.

الحل:

من البيانات المعطاة نجد أن $\bar{x} = 35, \quad s^2 = 4$

ومن متباينة تشيبشف نجد إن الفترة التي يقع فيها ما لا يقل عن $1 - \frac{1}{r^2}$ هي $(\bar{x} - rs, \bar{x} + rs)$ ونجد إن الفترة هي (32,38) فمن حدود الفترة يمكن تعيين قيمة r والتي تستخدم في تعيين النسبة.

$$\bar{x} - rs = 32 \Longrightarrow 35 - 2r = 32 \Longrightarrow 2r = 35 - 32$$

$$\Longrightarrow 2r = 3 \Longrightarrow r = \frac{3}{2} = 1.5$$

وبالتالي فإن النسبة هي:

$$1 - \frac{1}{r^2} = 1 - \frac{1}{(1.5)^2} = 1 - \frac{1}{2.25} = 1 - 0.4444 = 0.5556$$

مما سبق فإن ما لا يقل عن 55.6% من العمال بتلك الشركة يقع أعمارهم في الفترة (32,38) سنة.

٤-٨ الالتواء (Skewness)

الالتواء هو بعد التوزيع عن التماثل (lack of symmetry) كما سبق أن ذكرنا في الفصل الثاني شكل (٢-٩) فالتوزيع قد يكون متماثلاً أو ملتوياً جهة اليمين أو ملتوياً جهة اليسار.

- ففي حالة التوزيعات المتماثلة فإن:

الوسط الحسابي = الوسيط = المنوال

- وإذا كان التوزيع ملتوياً جهة اليمين فإن:

الوسط الحسابي > الوسيط > المنوال

- وإذا كان التوزيع ملتوياً جهة اليسار فإن:

المنوال > الوسيط > الوسط الحسابي

ويرمز للالتواء بالرمز sk ويمكن تعيين معامل الالتواء بعدة علاقات منها:

١- المقياس الأول: يحسب بقسمة الفرق بين الوسط الحسابي والمنوال على الانحراف المعياري فنحصل على

$$sk = \frac{\bar{x} - Mod}{s}$$

٢- المقياس الثاني: ويستخدم في حالة تعذر حساب المنوال فإن

$$sk = \frac{3(\bar{x} - Med)}{s}$$

٣- المقياس الثالث: ويمكن أيضا قياس الالتواء بدراسة المواقع النسبية للربيع الأول والوسيط والربيع الثالث للتوزيع التكراري. إذا كان التوزيع متماثلاً فإن الفرق بين الوسيط والربيع الأول يساوى الفرق بين الربيع الثالث والوسيط. وإذا كان التوزيع ملتوياً جهة اليمين فإن الربيع الأول يكون أقرب إلى الوسيط مقارنة بالربيع الثالث. وإذا كان التوزيع ملتوياً جهة اليسار فإن الربيع الثالث يكون أقرب إلى الوسيط مقارنة بالربيع الأول. ويكون مقياس الالتواء في هذه الحالة:

$$sk = \frac{(Q_3 - Med) - (Med - Q_1)}{(Q_3 - Med) + (Med - Q_1)}$$

ومن خلال قيمة معامل الالتواء يمكن تحديد نوع الالتواء كالتالي:

١- إذا كانت $sk = 0$ فإن المنحنى متماثل.

٢- إذا كانت $sk > 0$ فإن الالتواء موجب أي أن المنحنى ملتوٍ ناحية اليمين.

٣- إذا كانت $sk < 0$ فإن الالتواء سالب أي أن المنحنى ملتوٍ ناحية اليسار.

مثال (٤-١١)

إذا توفرت لدينا البيانات الآتية أحسب معامل الالتواء.

$$\bar{x} = 70,\ \ Med = 69, \qquad Mod = 68.5\,, Q_1 = 60, \qquad Q_3 = 80, s = 6$$

الحل

يمكن تعيين معامل الالتواء بأي من القوانين الثلاثة السابقة والتي ستعطى جميعها نفس نوع الالتواء لكنها قد تختلف في القيمة.

القانون الأول

$$sk = \frac{\bar{x} - Mod}{s}$$

$$= \frac{70 - 68.5}{6} = \frac{1.5}{6} = 0.25$$

القانون الثاني

$$sk = \frac{3(\bar{x} - Med)}{s}$$

$$= \frac{3(70 - 69.4)}{6} = \frac{1.8}{6} = 0.30$$

القانون الثالث

$$sk = \frac{(Q_3 - Med) - (Med - Q_1)}{(Q_3 - Med) + (Med - Q_1)}$$

$$= \frac{(80 - 69.4) - (69.4 - 60)}{(80 - 69.4) + (69.4 - 60)} = \frac{1.2}{20} = 0.06$$

ومن هذه النتائج يتضح أن الالتواء موجب أي أن المنحني ملتوٍ ناحية اليمين.

مثال (٤-١٢)

أوجد معامل الالتواء للبيانات التالية مع تحديد نوع الالتواء: 41, 32, 46, 28, 36, 20, 35, 23

الحل:

من الملاحظ أن البيانات لا يوجد لها منوال. لذا لن نستطيع استخدام القانون الأول، لذا سوف نلجأ لاستخدام القانون الثاني والثالث.

١- حساب الالتواء باستخدام العلاقة:

$$sk = \frac{3(\bar{x} - Med)}{s}$$

ولحساب الوسط الحسابي والانحراف المعياري نكون الجدول التالي:

x	x^2
41	1681
32	1024
46	2116
28	784
36	1296
20	400
35	1225
23	529
$\sum x = 261$	$\sum x^2 = 9055$

الوسط الحسابي:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{261}{8} = 32.625$$

الانحراف المعياري:

$$s = \sqrt{\frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2\right]}$$

$$= \sqrt{\frac{1}{8-1}\left[9055 - 8\left(\frac{261}{8}\right)^2\right]}$$

$$= \sqrt{77.125} = 8.7821$$

الوسيط للبيانات:

نقوم أولاً بترتيب البيانات تصاعدياً:

القيمة	الترتيب
20	1
23	2
28	3
32	4
35	5
36	6
41	7
46	8

ثانياً نحدد عدد البيانات؛ زوجي أم فردى.

عدد العناصر $n = 8$ وهو عدد زوجي لذا فإنه توجد قيمتان في الوسط ولهما الترتيب

$$\frac{n}{2} = \frac{8}{2} = 4, \ \frac{n}{2} + 1 = 4 + 1 = 5$$

فإن الوسيط هو

$$Med = \frac{X_{(4)} + X_{(5)}}{2} = \frac{32 + 35}{2} = 33.5$$

لاحظ أن:

$$\bar{x} < Med$$

فإن المنحنى (التوزيع) ملتوٍ ناحية اليسار.

ومعامل الالتواء هو

$$sk = \frac{3(\bar{x} - Med)}{s}$$

$$= \frac{3(32.62 - 33.5)}{8.7821} = \frac{-2.64}{8.7821} = -0.3006$$

قيمة معامل الالتواء سالبه وبالتالي فإن التوزيع ملتوٍ ناحية اليسار.

٢- حساب الالتواء باستخدام العلاقة:

$$sk = \frac{(Q_3 - Med) - (Med - Q_1)}{(Q_3 - Med) + (Med - Q_1)}$$

يجب تعيين كلاً من الربيع الأول والثالث كالتالي:

رتبة الربيع الأول هو

$$\frac{n}{4} = \frac{8}{4} = 2$$

أي أن الربيع الأول هو القيمة الثانية في البيانات المرتبة

$$Q_1 = X_{(2)} = 23$$

رتبة الربيع الثالث هو

$$\frac{3n}{4} = \frac{3 \times 8}{4} = \frac{24}{4} = 6$$

أي أن الربيع الثالث هو القيمة السادسة في البيانات المرتبة

$$Q_3 = X_{(6)} = 36$$

فإن معامل الالتواء

$$sk = \frac{(Q_3 - Med) - (Med - Q_1)}{(Q_3 - Med) + (Med - Q_1)}$$

$$= \frac{(36 - 33.5) - (33.5 - 23)}{(36 - 33.5) + (33.5 - 23)} = -\frac{8}{13} = -0.6154$$

إذاً قيمة معامل الالتواء سالبة وبالتالي فإن التوزيع ملتو ناحية اليسار.

مثال (٤-١٣)

البيانات التالية تمثل الأجور الأسبوعية لمائة عامل مبوبة في الجدول التالي:

فترات الأجور	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64
عدد العمال	10	20	40	20	10

احسب قيمة معامل الالتواء وبين نوع التوزيع؟

الحل:

من المعلوم أن معامل الالتواء يمكن حسابه باستخدام أي من القوانين الثلاثة واستخدام واحدة من تلك العلاقات يكفي لتعيين معامل الالتواء، لذا فإننا هنا سنقوم باستخدام العلاقة الأولى.

المنوال:

الفترات التقريبية	الفترات الفعلية	f
40 – 44	39.5 – 44.5	10
45 – 49	44.5 – 49.5	20
50 – 54	49.5 – 54.5	40
55 – 59	54.5 – 59.5	20
60 – 64	59.5 – 64.5	10

من الجدول التكراري نجد إن أكبر تكرار هو 40 وبالتالي فإن الفترة التي تحتوي المنوال هي 54.5 – 49.5 ومنها نجد أن التكرار السابق لأكبر تكرار هو 20 والتكرار اللاحق لأكبر تكرار هو 20

$$f = 40, \quad f_1 = 20, \quad f_2 = 20, \quad L = 49.5, \quad h = 5$$
$$\Longrightarrow \Delta_1 = f - f_1 = 40 - 20 = 20, \qquad \Delta_2 = f - f_2 = 40 - 20 = 20$$

ويمكن تعيين المنوال من القانون

$$mod = L + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times h$$
$$= 49.5 + \left(\frac{20}{20 + 20}\right) \times 5 = 52$$

الوسط الحسابي:

فترات الأجر	f_i	x_i	$x_i f_i$
39.5 – 44.5	10	42	420
44.5 – 49.5	20	47	940
49.5 – 54.5	40	52	2080
54.5 – 59.5	20	57	1140
59.5 – 64.5	10	62	620
المجموع	100		5200

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i = \frac{1}{100}(5200) = 52$$

ويمكن أيضا حساب الوسيط كالتالي:

نحسب التكرار المتجمع الصاعد

الحدود العليا للفترات	التكرار المتجمع الصاعد
أقل من 39.5	0
أقل من 44.5	10
أقل من 49.5	30
أقل من 54.5	70
أقل من 59.5	90
أقل من 64.5	100

رتبة الوسيط هي

$$\frac{n}{2} = \frac{100}{2} = 50$$

من التكرار المتجمع الصاعد فإن الفترة التي تحتوى الوسيط هي 54.5 – 49.5، ومنها فإن

$$L = 49.5, \quad h = 5, \; F_1 = 30, \; F_2 = 70$$

والوسيط هو

$$Med = L + \left(\frac{\frac{n}{2} - F_1}{F_2 - F_1}\right) \times h$$

$$= 49.5 + \left(\frac{50 - 30}{70 - 30}\right) \times 5 = 52$$

ونجد أن

$$\bar{x} = Mod = Med = 52$$

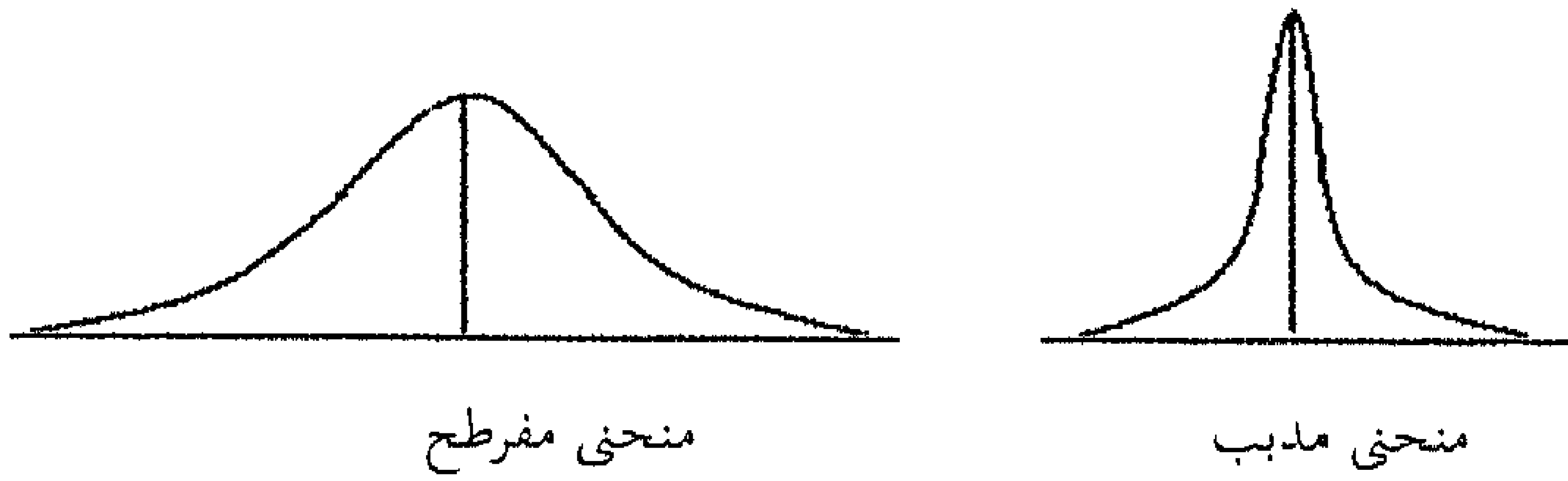
لذا فإن توزيع الأجور متماثل وبالتالي سيكون معامل الالتواء مساوياً للصفر.

$$sk = \frac{\bar{x} - Mod}{s} = \frac{52 - 52}{s} = \frac{0}{s} = 0$$

$$sk = \frac{3(\bar{x} - Med)}{s} = \frac{3(52 - 52)}{s} = 0$$

٤-٩ التفرطح (Kurtosis)

عند تمثيل التوزيع التكراري في شكل منحنى تكراري، قد يكون هذا المنحنى منبسطاً ، أو مدبباً، فعندما يتركز عدد أكبر من القيم(المشاهدات) بالقرب من منتصف المنحنى، ويقل في طرفيه، يكون المنحنى مدبباً، وعندما يتركز عدد أكبر على طرفي المنحنى، ويقل بالقرب من المنتصف يكون المنحنى مفرطحاً، أو منبسطا، ويظهر ذلك من الشكل التالي:



ويمكن قياس التفرطح باستخدام عدد من الطرق، ومنها طريقة العزوم، حيث يحسب معامل التفرطح (K) بتطبيق المعادلة التالية :

$$k = \frac{m_4}{s^4}$$

حيث s هي الانحراف المعياري، m_4 هو العزم الرابع للبيانات حول وسطها الحسابي. حيث إن العزم الرائي للبيانات حول الوسط يعطى بالعلاقة التالية:

- بالنسبة للبيانات المفردة

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r, \qquad r = 1, 2, 3, \ldots$$

- بالنسبة للبيانات المبوبة

$$m_r = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^r f_i, \quad r = 1, 2, 3, \ldots$$

ومعامل التفرطح للبيانات المتماثلة والتي تتوزع توزيعاً طبيعياً يساوي 3، ومن ثم يمكن وصف منحنى

التوزيع من حيث التفرطح ، والمدبب كما يلي:

- إذا كان $k = 3$ أو $k - 3 = 0$ فإن منحنى التوزيع معتدلاً.
- إذا كان $k > 3$ أو $k - 3 > 0$ فإن منحنى التوزيع مدبباً.
- إذا كان $k < 3$ أو $k - 3 < 0$ فإن منحنى التوزيع منبسطاً (مفرطحاً).

مثال (٤-١٤)

أوجد معامل التفرطح للبيانات التالية مع تحديد نوع المنحنى من حيث التفرطح ,66, 85, 52, 78, 80
91, 74, 58

الحل:

معامل التفرطح يعطي بالعلاقة:

$$k = \frac{m_4}{s^4}$$

الوسط الحسابي

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{8}(66 + 85 + 52 + 78 + 80 + 91 + 74 + 58)$$

$$= \frac{584}{8} = 73$$

نقوم بتكوين الجدول التالي لحساب التباين والعزم الرابع:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^4$
66	-7	49	2401
85	12	144	20736
52	-21	441	194481
78	5	25	625
80	7	49	2401
91	18	324	104976
74	1	1	1
58	-15	225	50625
$\sum x = 584$	0	$\sum(x - \bar{x})^2 = 1258$	$\sum(x - \bar{x})^4 = 376246$

ومن البيانات أعلاه نجد أن الانحراف المعياري هو:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1258}{8-1}} = 13.4058$$

والعزم الرابع للبيانات هو:

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 = \frac{376246}{8} = 47030.75$$

فإن معامل التفرطح هو:

$$k = \frac{m_4}{s^4} = \frac{47030.75}{(13.4058)^4} = 1.4562,$$

$$k - 3 = 1.4562 - 3 = -1.5438 < 0$$

وحيث أن $k - 3 < 0$ فإن توزيع البيانات مفرطح.

مثال (٤-١٥)

البيانات التالية تمثل أوزان مائة طالب مبوبة في الجدول التالي:

فترات الأوزان	40 –	45 –	50 –	55 –	60 –	65 –
عدد الطلاب	7	18	40	20	10	5

احسب قيمة معامل التفرطح وبين نوع منحنى التوزيع من حيث التفرطح؟

الحل:

أولاً: حساب المتوسط الحسابي للبيانات:

الفترات	f	x	xf
40 – 45	7	42.5	297.5
45 – 50	18	47.5	855
50 – 55	40	52.5	2100
55 – 60	20	57.5	1150
60 – 65	10	62.5	625
65 – 70	5	67.5	337.5
المجموع	100		5365

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i = \frac{5365}{100} = 53.65$$

ثانياً: حساب الانحراف المعياري والعزم الرابع حول الوسط:

الفترات	f	x	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^4$	$(x_i - \bar{x})^2 f_i$	$(x_i - \bar{x})^4 f_i$
40 – 45	7	42.5	-11.15	124.3225	15456.08	870.2575	108192.6
45 – 50	18	47.5	-6.15	37.8225	1430.542	680.805	25749.75
50 – 55	40	52.5	-1.15	1.3225	1.749006	052.900	69.96025
55 – 60	20	57.5	3.85	14.8225	219.7065	296.450	4394.13
60 – 65	10	62.5	8.85	78.3225	6134.414	783.225	61344.14
65 – 70	5	67.5	13.85	191.8225	36795.87	959.1125	183979.4
المجموع	100					3642.75	383729.9

من الجدول السابق نجد أن الانحراف المعياري هو:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}$$

$$= \sqrt{\frac{3642.75}{100-1}} = 36.7949$$

العزم الرابع حول المتوسط هو:

$$m_4 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^4 f_i$$

$$= \frac{383729.9}{100} = 3837.299$$

فإن معامل التفرطح هو:

$$k = \frac{m_4}{s^4} = \frac{3837.299}{(36.7949)^4} = 0.0021$$

$$k - 3 = 0.0021 - 3 = -2.9979$$

وحيث إن $k < 3$ وأيضا $k - 3 < 0$ فإن توزيع البيانات مفرطح.

٤-١٠ تطبيقات باستخدام برنامج SPSS

تطبيق (٤-١)

من مثال (٤-١): فيما يلي التوزيع التكراري لإنتاج مجموعات من قطع الأرض الزراعية المتساوية من أحد المحاصيل (بآلاف الكيلوجرام):

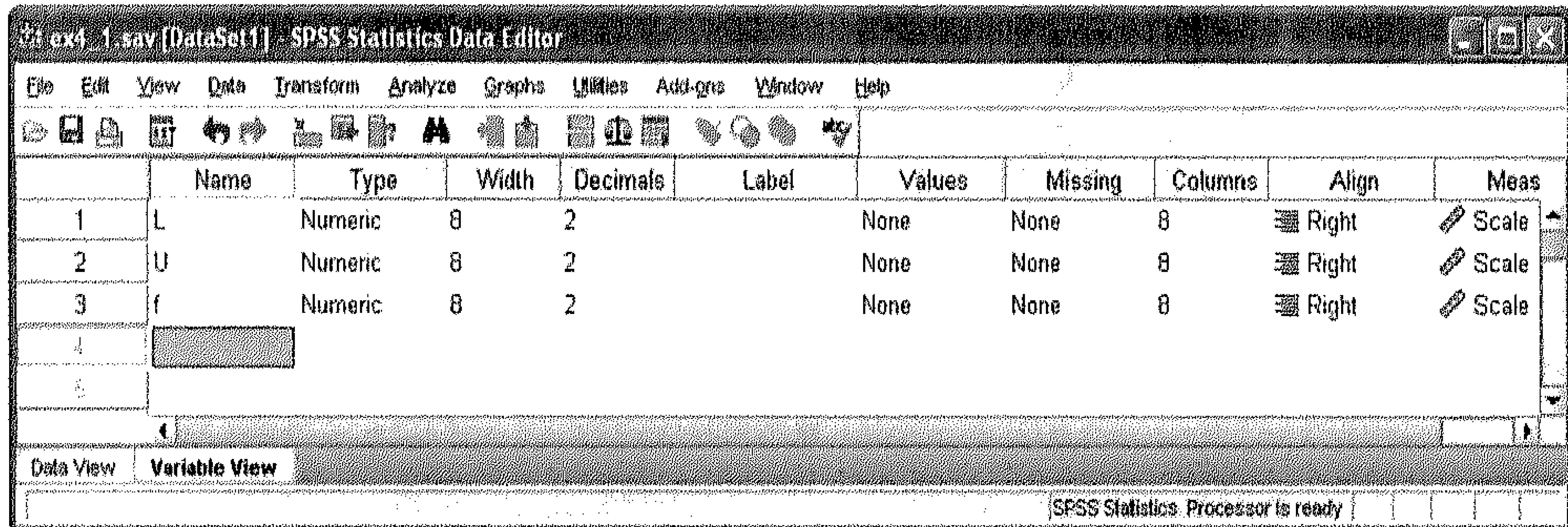
فئات الإنتاج	26–	31–	36–	41–	46–	51–	56–	61–
عدد القطع	4	5	23	58	61	30	4	3

المطلوب إيجاد المدى للإنتاج.

الحل:

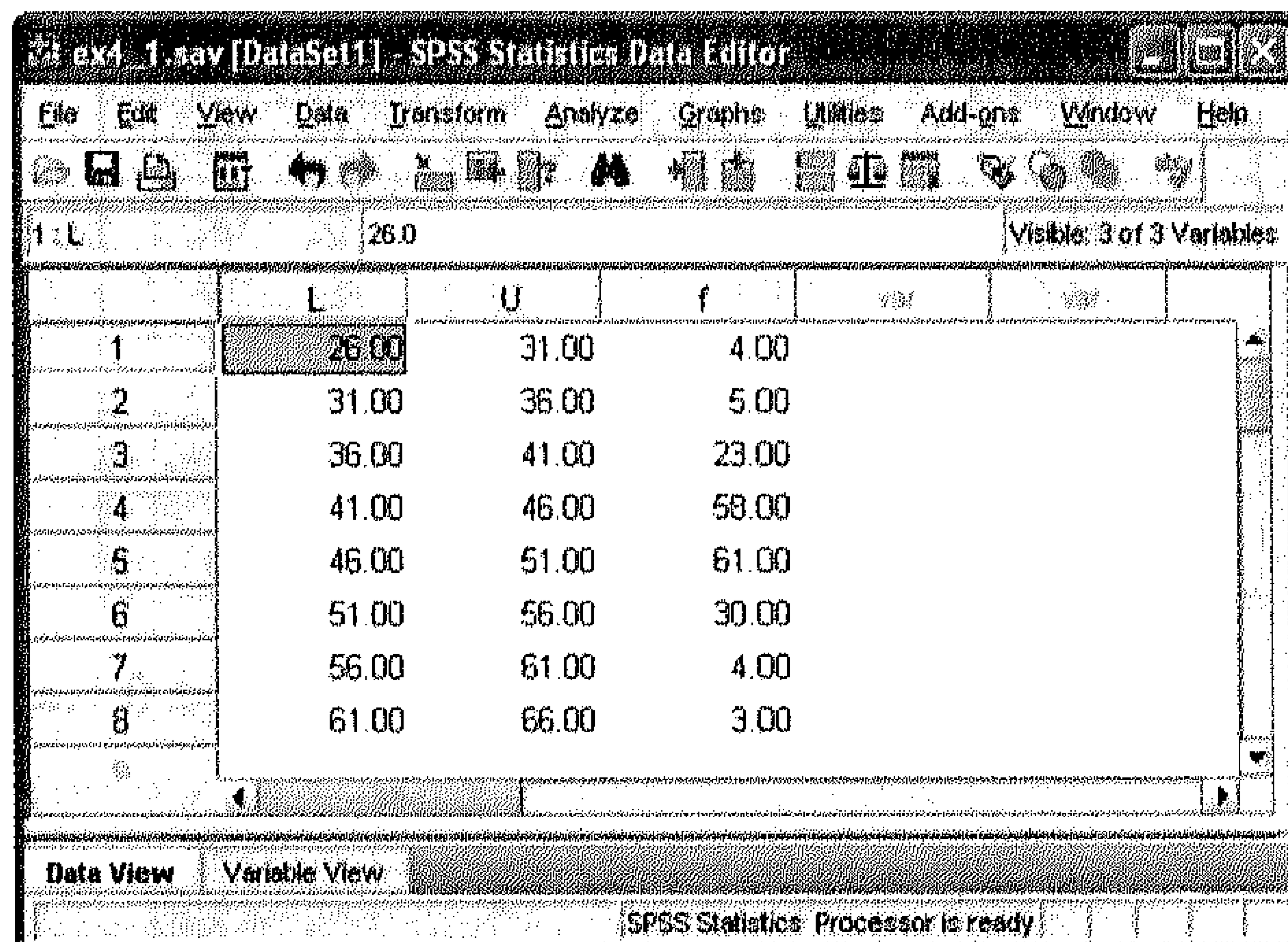
أولا: إدخال البيانات للبرنامج وذلك باتباع الخطوات التالية:

١- نقوم بتعريف ثلاث متغيرات L, U, f في نافذة Variable View



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Meas
1	L	Numeric	8	2		None	None	8	Right	Scale
2	U	Numeric	8	2		None	None	8	Right	Scale
3	f	Numeric	8	2		None	None	8	Right	Scale

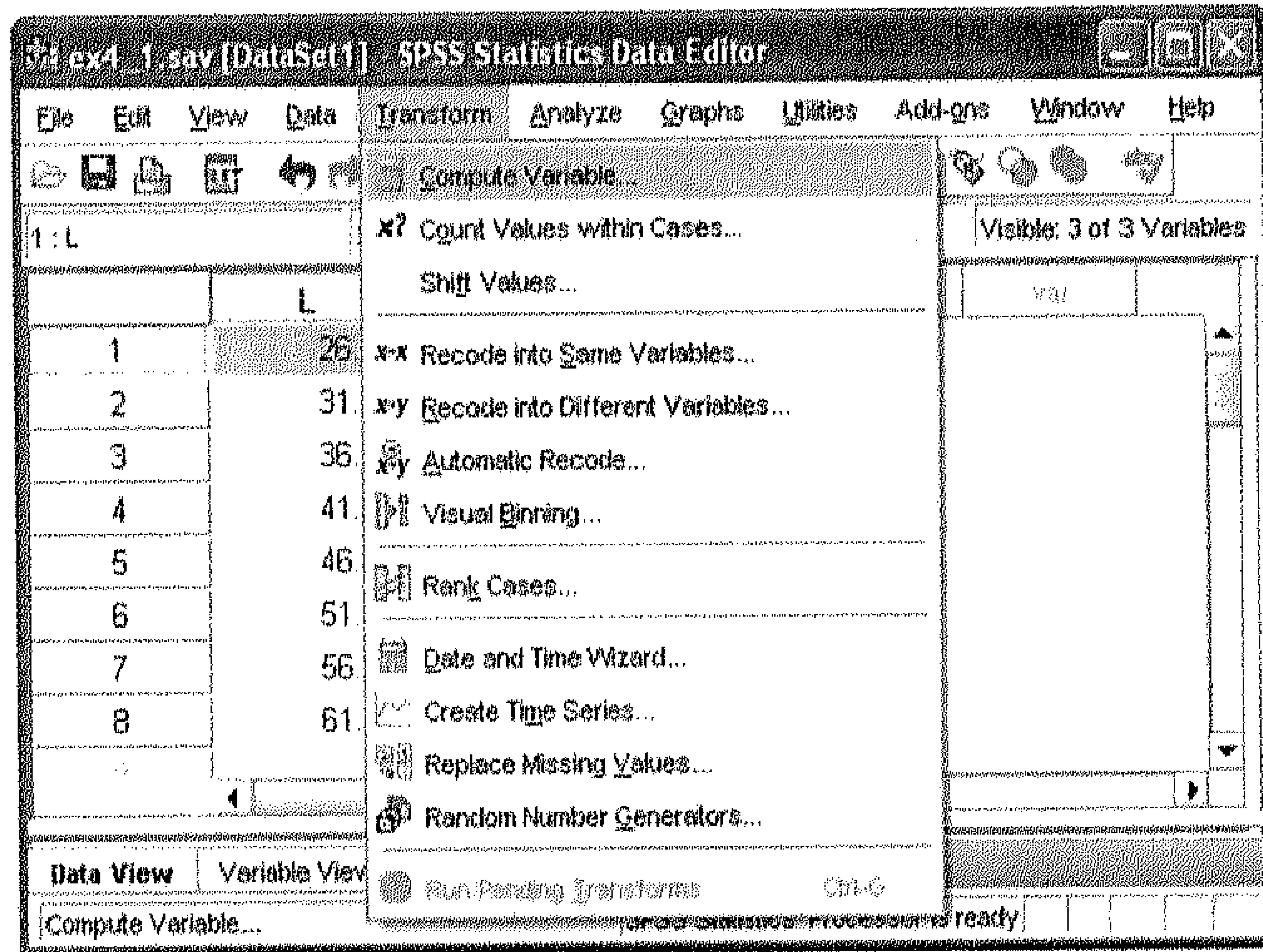
٢- من نافذة Data View نقوم بإدخال البيانات الخاصة ببداية الفترات L ونهاية الفترات U والتكرار f



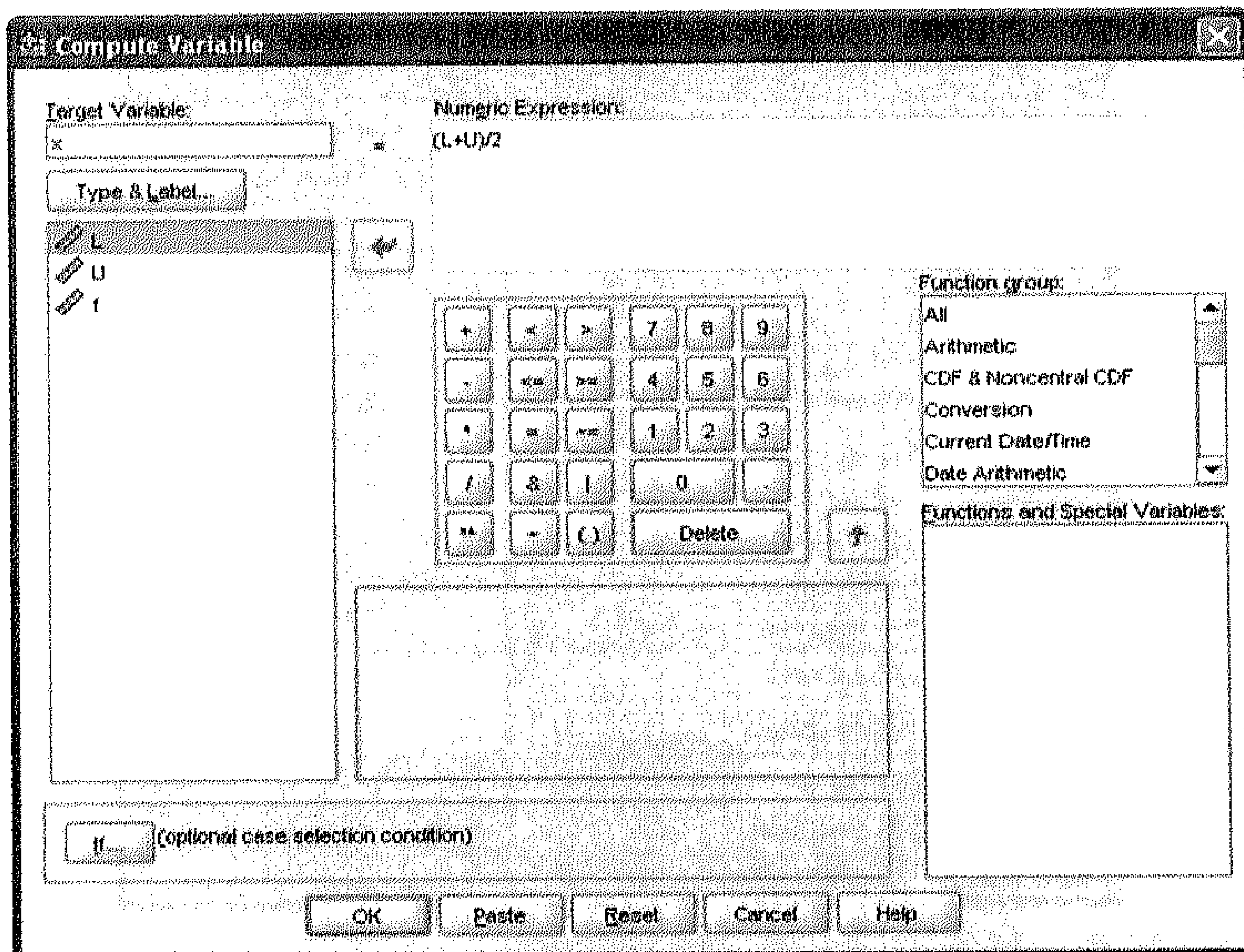
	L	U	f
1	26.00	31.00	4.00
2	31.00	36.00	5.00
3	36.00	41.00	23.00
4	41.00	46.00	58.00
5	46.00	51.00	61.00
6	51.00	56.00	30.00
7	56.00	61.00	4.00
8	61.00	66.00	3.00

ثانياً: تعيين مراكز الفترات وذلك بإتباع الخطوات التالية:

١- من قائمة Transform نختار الأمر Compute Variable



٢- تظهر نافذة جديدة نكتب X في شريط Target Variable ونكتب العلاقة التي سيحسب بها مركز الفترات في خانة Numeric Expression



٣- ثم نضغط على Ok فنعود لملف البيانات وقد أضاف متغيراً جديداً X يحتوى على مركز الفترات

ex4_1.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : L 26.0 Visible: 4 of 4 Variables

	L	U	f	x	var
1	26.00	31.00	4.00	28.50	
2	31.00	36.00	5.00	33.50	
3	36.00	41.00	23.00	38.50	
4	41.00	46.00	58.00	43.50	
5	46.00	51.00	61.00	48.50	
6	51.00	56.00	30.00	53.50	
7	56.00	61.00	4.00	58.50	
8	61.00	66.00	3.00	63.50	

Data View Variable View

SPSS Statistics Processor is ready

ثالثاً: تحديد أن f هي التكرار وذلك باتباع الخطوات التالية:

١- من قائمة Data نختار weight cases لتحديد أن f هي تكرار للمتغير x

ex4_1.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

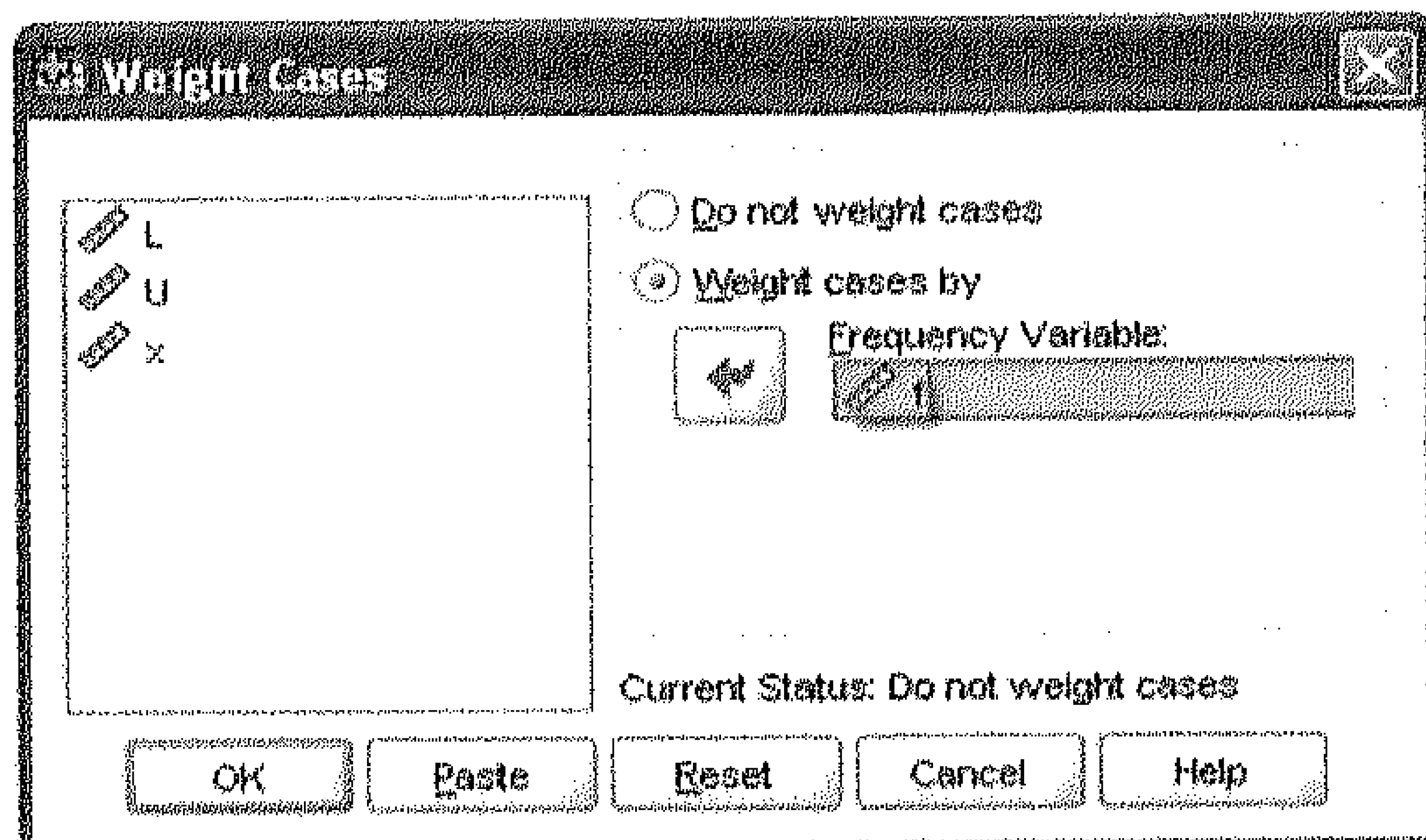
Define Variable Properties...
Copy Data Properties...
New Custom Attribute...
Define Dates...
Define Multiple Response Sets...
Identify Duplicate Cases...
Sort Cases...
Sort Variables...
Transpose...
Restructure...
Merge Files
Aggregate...
Copy Dataset
Split File...
Select Cases...
Weight Cases...

Visible: 4 of 4 Variables

	x
1	28.50
2	33.50
3	38.50
4	43.50
5	48.50
6	53.50
7	58.50
8	63.50

Data View
Weight Cases... Processor is ready

٢- نحدد الاختيار weight case by ثم ننقل المتغير f لخانة frequency variable

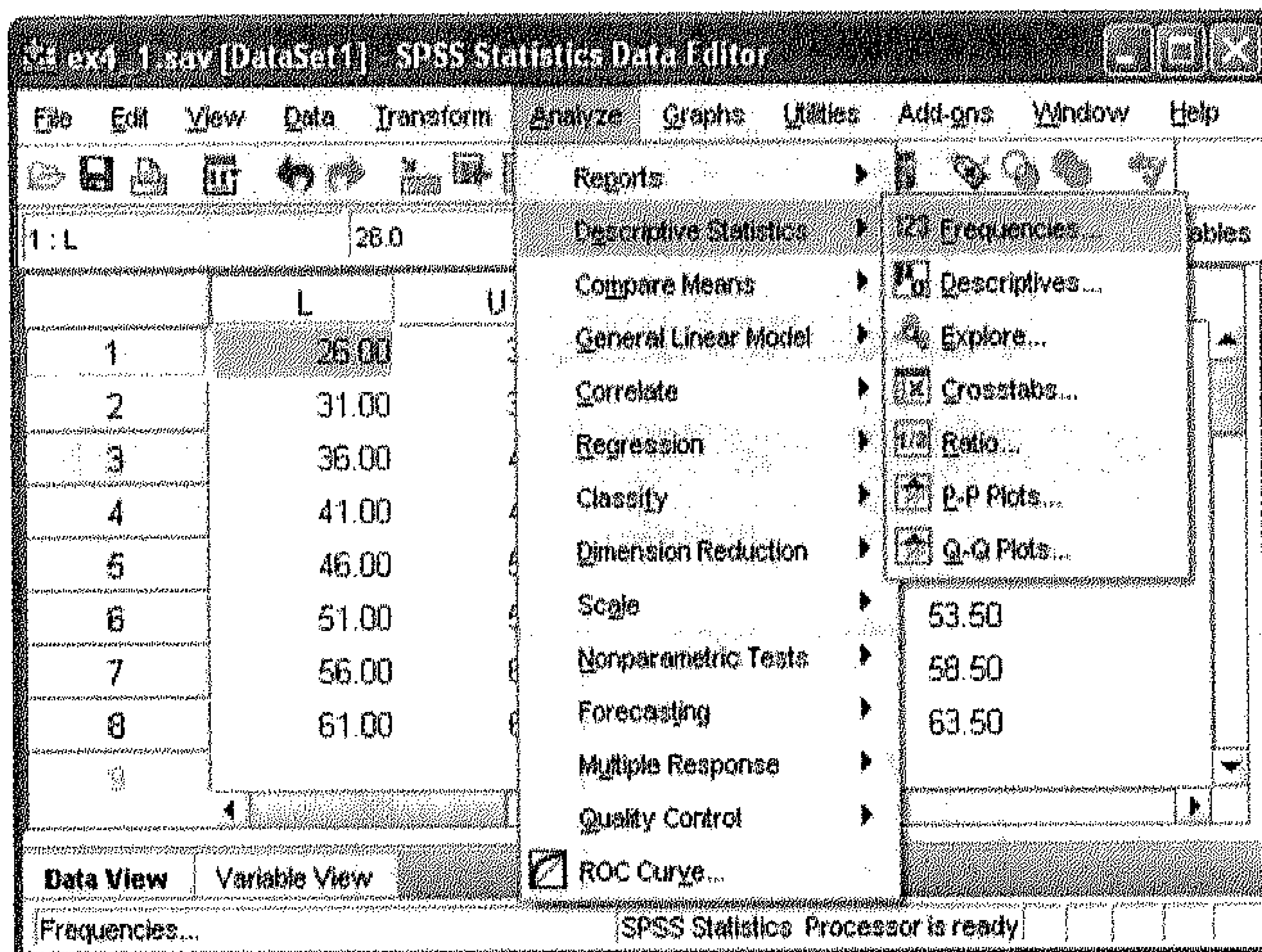


٣- نضغط على Ok فنعود لملف البيانات.

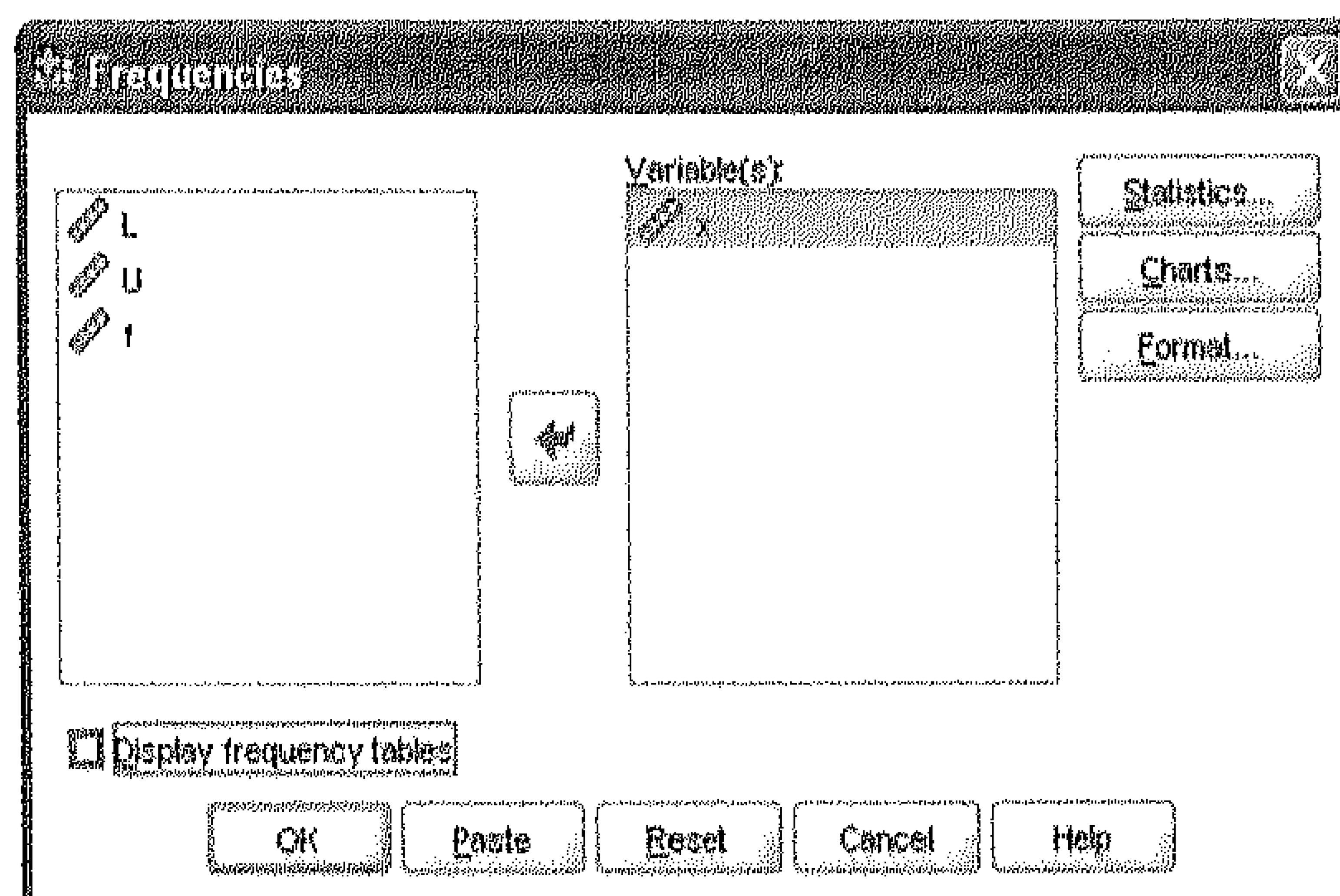
رابعاً: تعيين المقاييس المطلوبة وذلك باتباع الخطوات التالية:

١- من قائمة Analyze نختار Descriptive Statistics

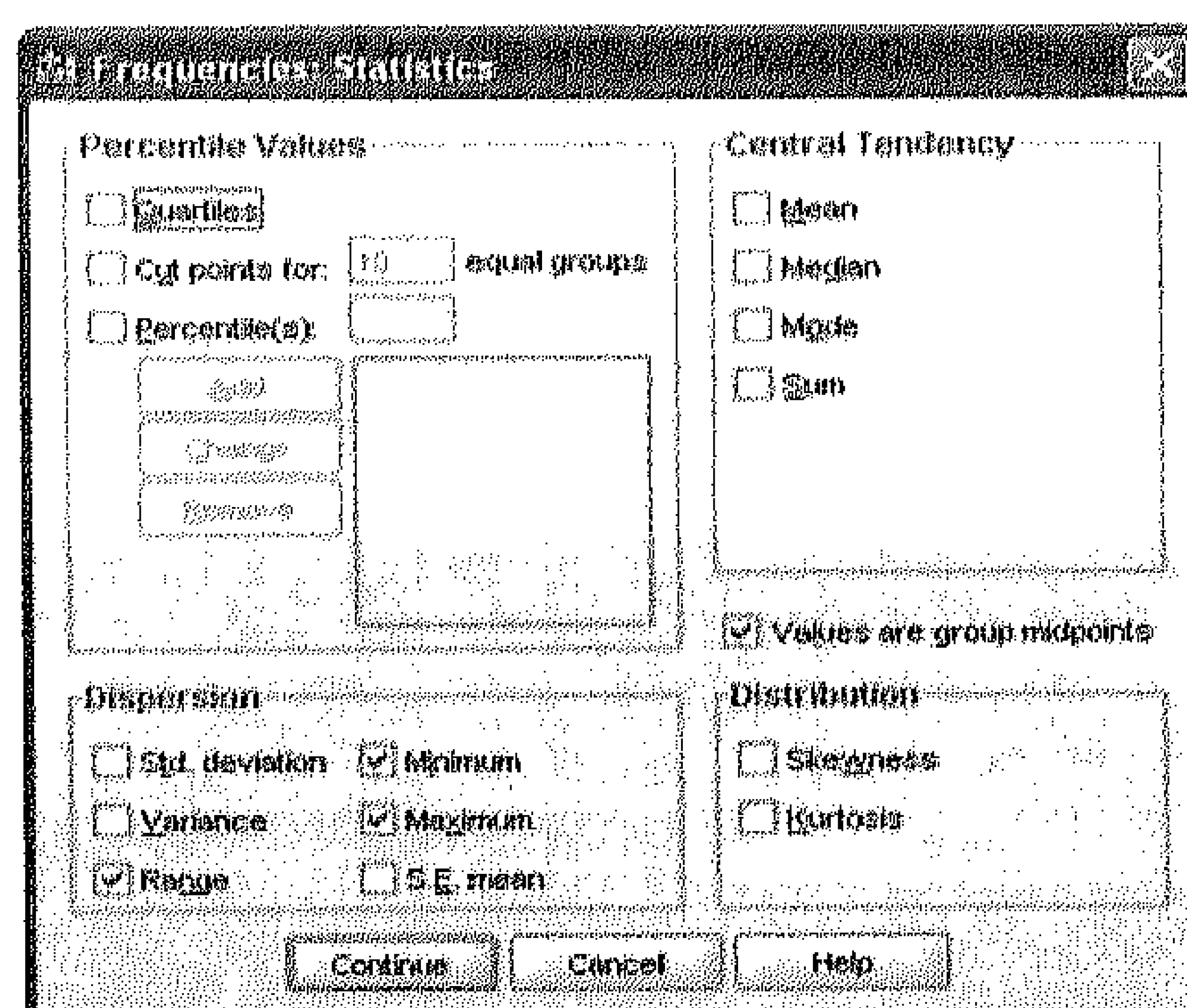
٢- ومن القائمة المنسدلة نختار الأمر Frequencies



٣- تظهر نافذة جديدة ننقل المتغير X لقائمة Variables ثم نختار Statistics



٤- من نافذة Frequencies: Statistics نختار من قائمة Dispersion الأمر Range وأيضا نحدد Maximum, Minimum



٥- ثم نضغط على Continue فنعود للشاشة السابقة فنضغط على Ok فتظهر النتائج كما يلي:

Statistics

x

N	Valid	8
	Missing	0
Range		35.00
Minimum		28.50
Maximum		63.50

من الجدول السابق نجد أن أقل وأكبر قيمة في البيانات هي $X_{min} = 28,\ X_{max} = 63,$ ومدى البيانات $Range = 35$ ومن المعلوم أن المدى يعين من العلاقة

$$Range = X_{max} - X_{min} = 63 - 28 = 35$$

تطبيق (٤-٢):

من مثال (٤-٣) اختيرت عينة من 10 مكالمات دولية بعد الساعة 11:00 مساءً وقبل الساعة 8:00 صباحاً فكانت أطوالها بالدقائق هي: ,15 ,12 ,6 ,20 ,10 ,13 ,3 ,9 ,4 ,8 احسب

(أ) المدى لأطوال المكالمات.

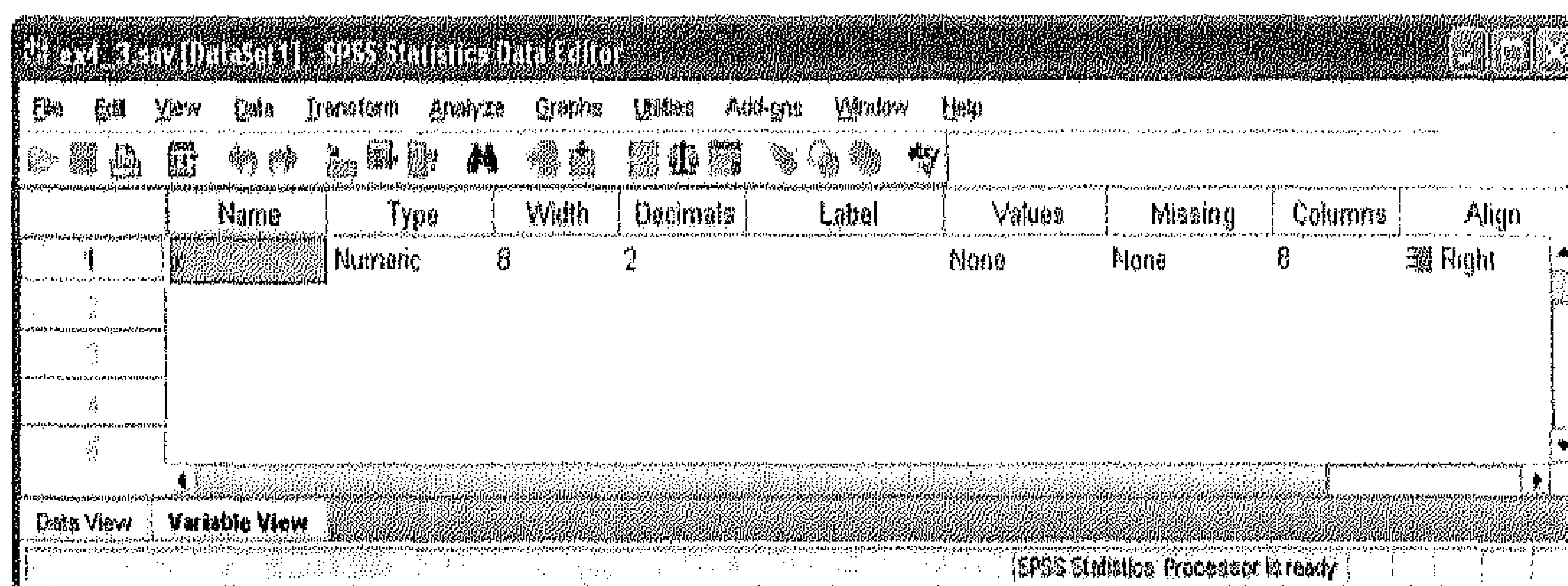
(ب) الانحراف المعياري والتباين لأطوال المكالمات.

(ج) نصف المدى الربيعي للبيانات.

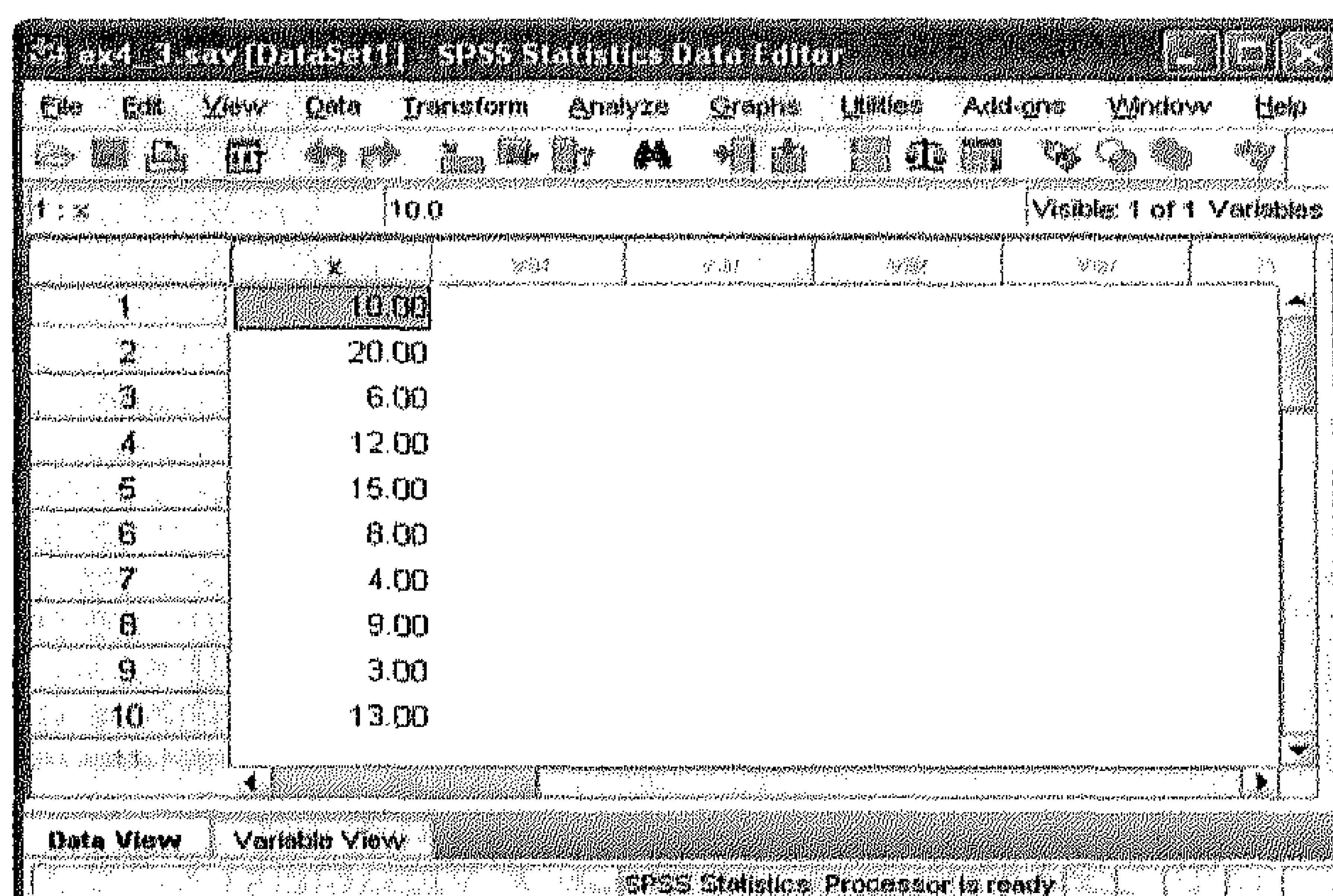
الحل

أولاً: نقوم بإدخال البيانات للبرنامج وذلك باتباع الخطوات التالية:

١- نقوم أولاً بتعريف متغير X في نافذة Variable View

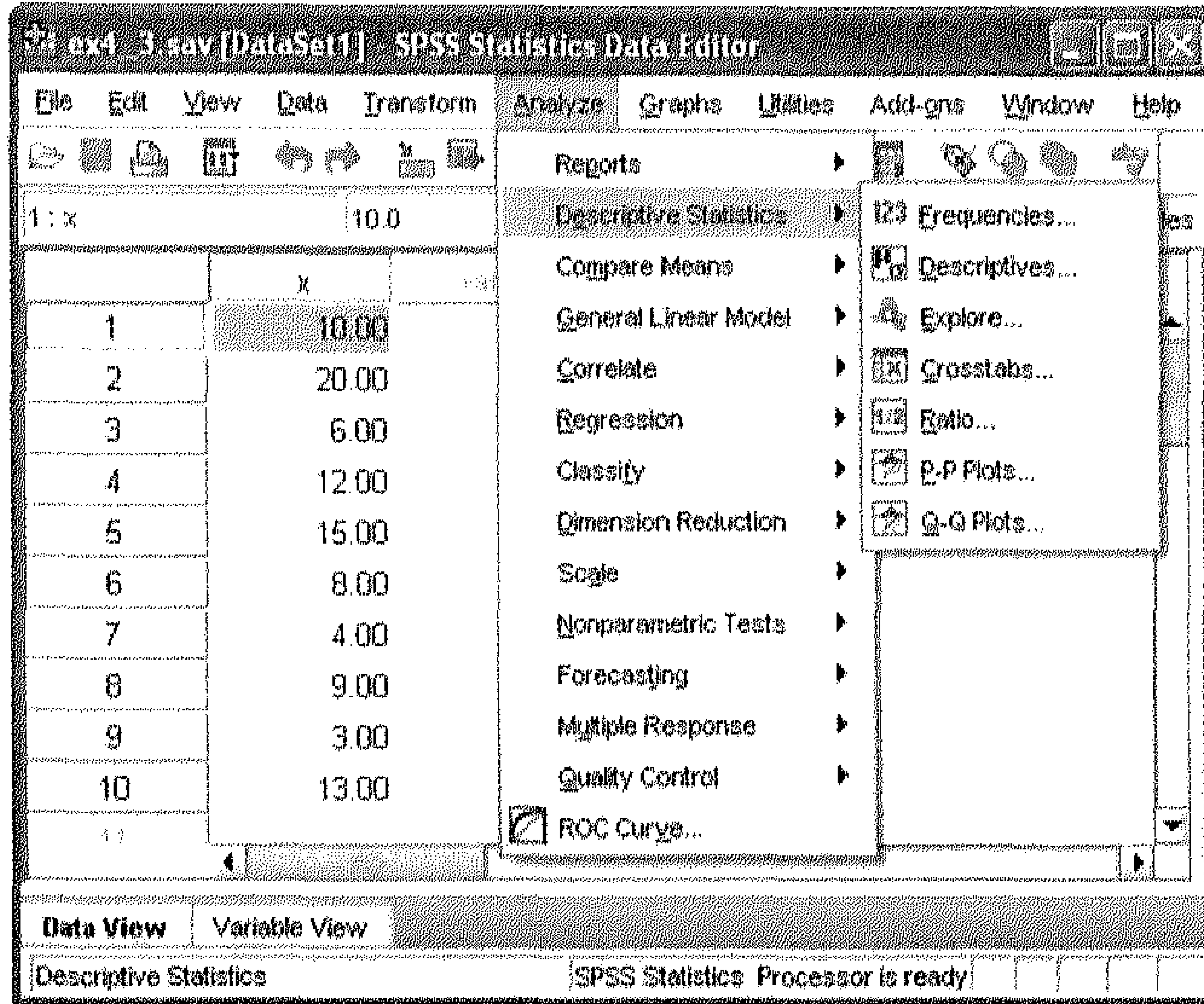


٢- ثم ننتقل لنافذة Data View ونقوم بإدخال البيانات.



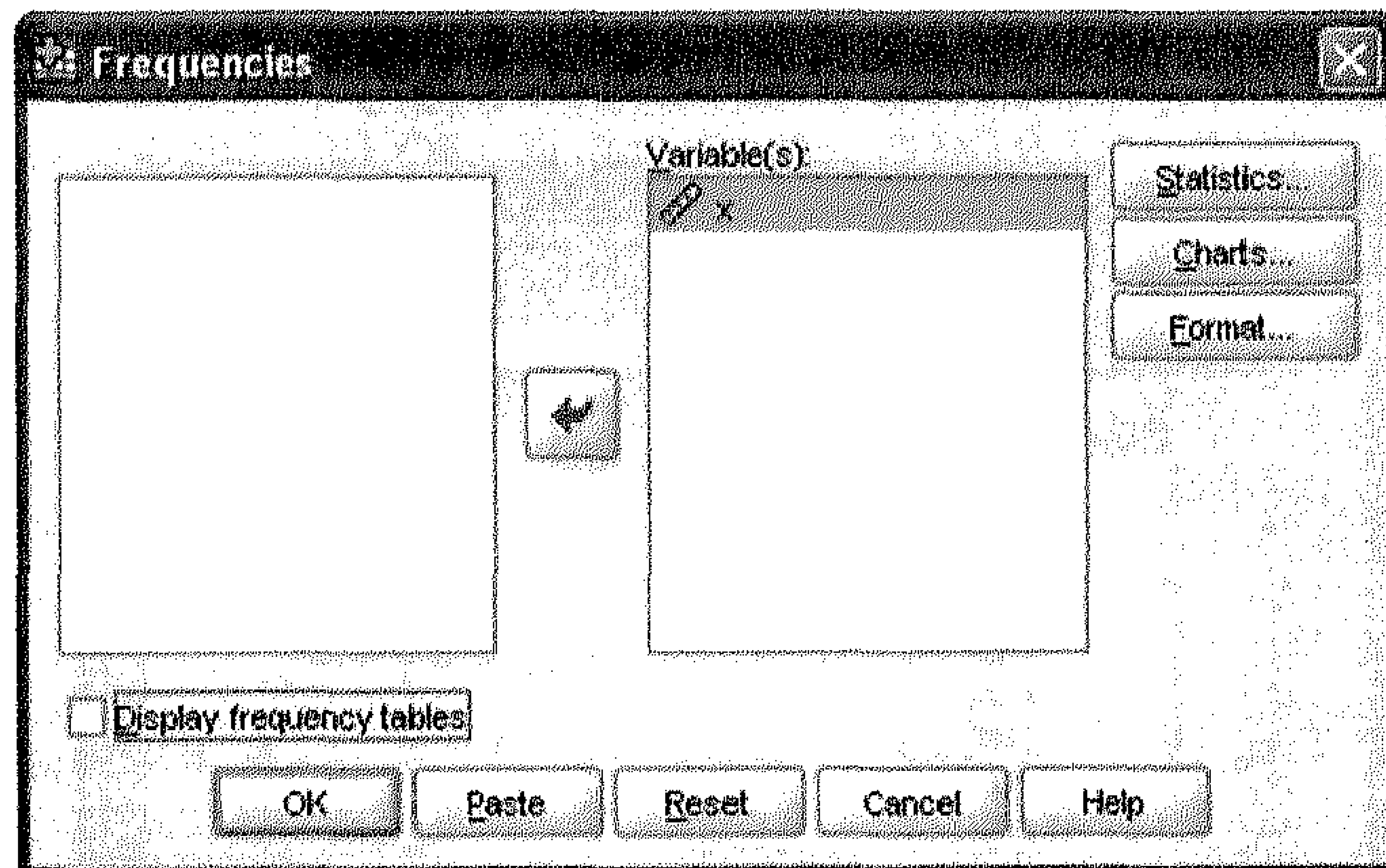
ثانياً: حساب المقاييس المطلوبة وذلك باتباع الخطوات التالية:

١- من قائمة Analyze نختار Descriptive Statistics فتظهر قائمة منسدلة سوف نستخدم بعضها:

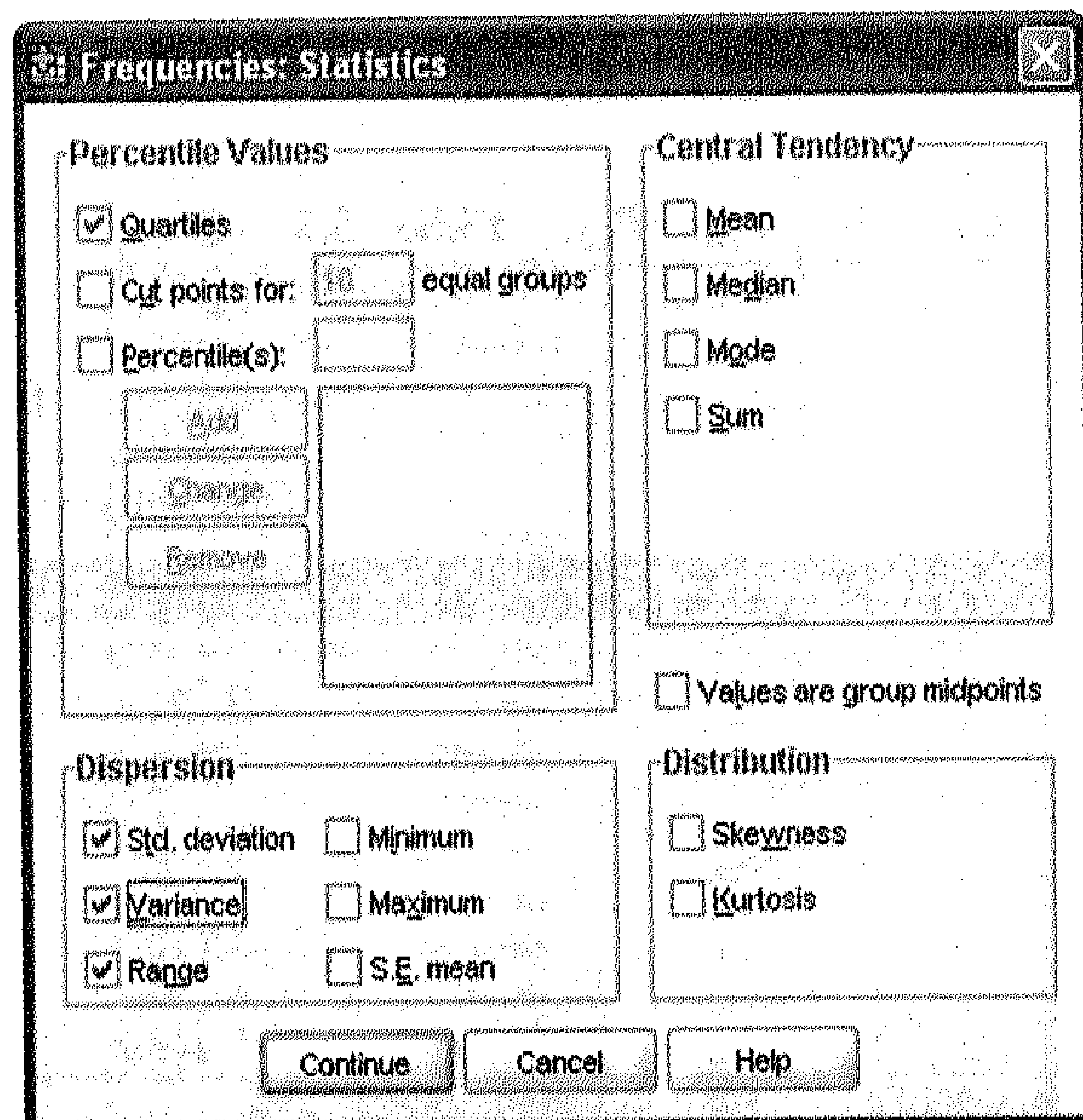


(أ) الأمر Frequencies

- بالضغط على الأمر Frequencies تظهر نافذة جديدة بعنوان Frequencies



- ننقل المتغير X لقائمة Variable(s) ثم نضغط على Statistics فتظهر نافذة جديدة



- نختار Quartiles من قائمة Percentiles ومن قائمة Dispersion نختار Std. Deviation لحساب الانحراف المعياري، Variance لحساب التباين، Range لحساب المدى،
- ثم نضغط على Continue فنعود للشاشة السابقة ثم نضغط على Ok فتظهر النتائج التالية

Statistics

x

N	Valid	10
	Missing	0
Std. Deviation		5.20683
Variance		27.111
Range		17.00
Percentiles	25	5.5000
	50	9.5000
	75	13.5000

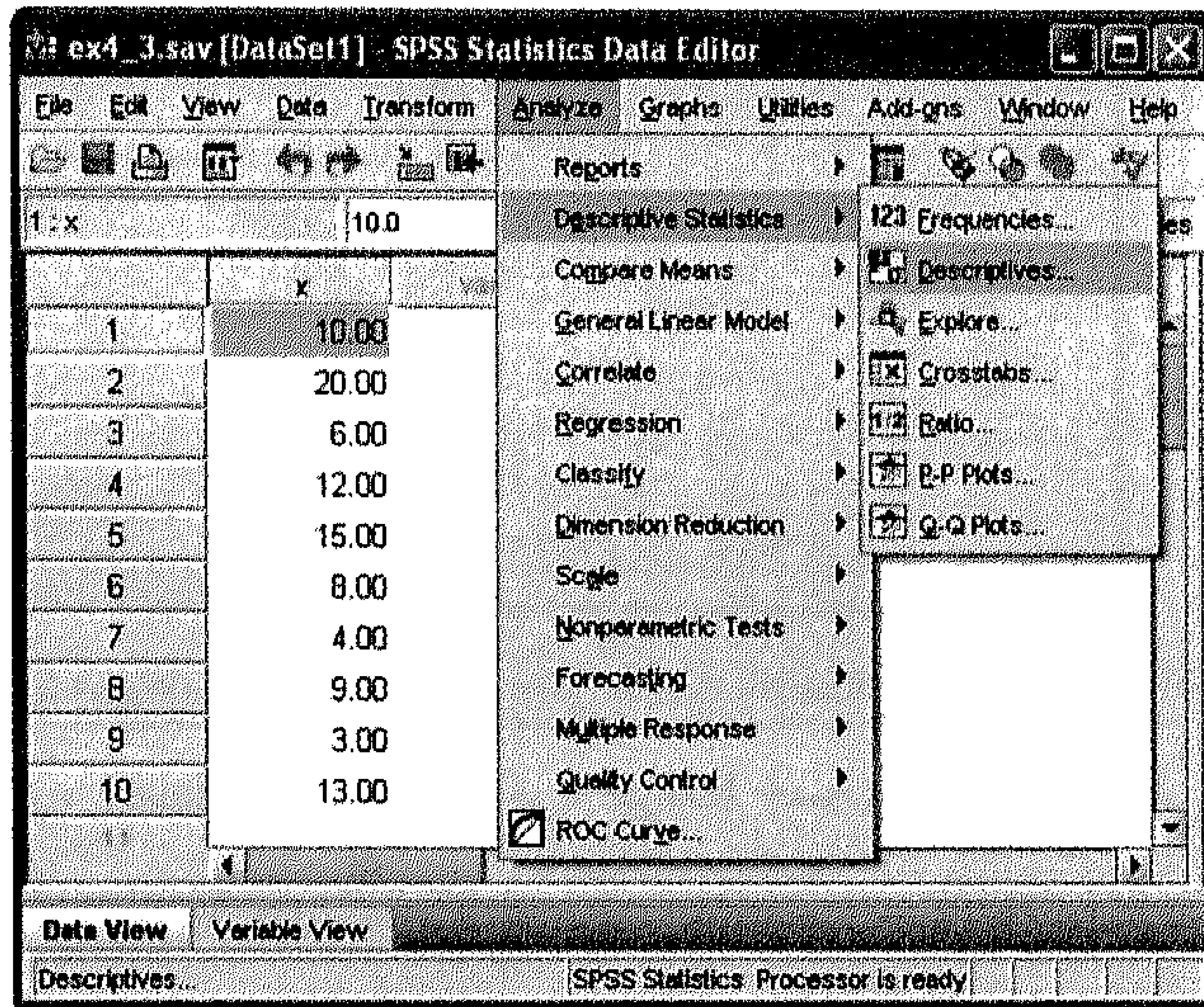
نجد من الجدول السابق أن: عدد القيم $n = 10$، الانحراف المعياري $s = std.Deviation = 5.20683$ والتباين $s^2 = Variance = 27.111$ والمدى للبيانات $R = Range = 17$، المئين الخامس والعشرون والذى يساوى الربيع الاول هو $P_{25} = 5.5$، المئين الخمسون والذى يساوى الربيع الثاني وأيضاً الوسيط هو $P_{50} = 9.5$

المئين الخامس والسبعين والذي يساوى الربيع الثالث هو $P_{75} = 13.5$ فإن نصف المدى الربيعي يمكن حسابه من العلاقة

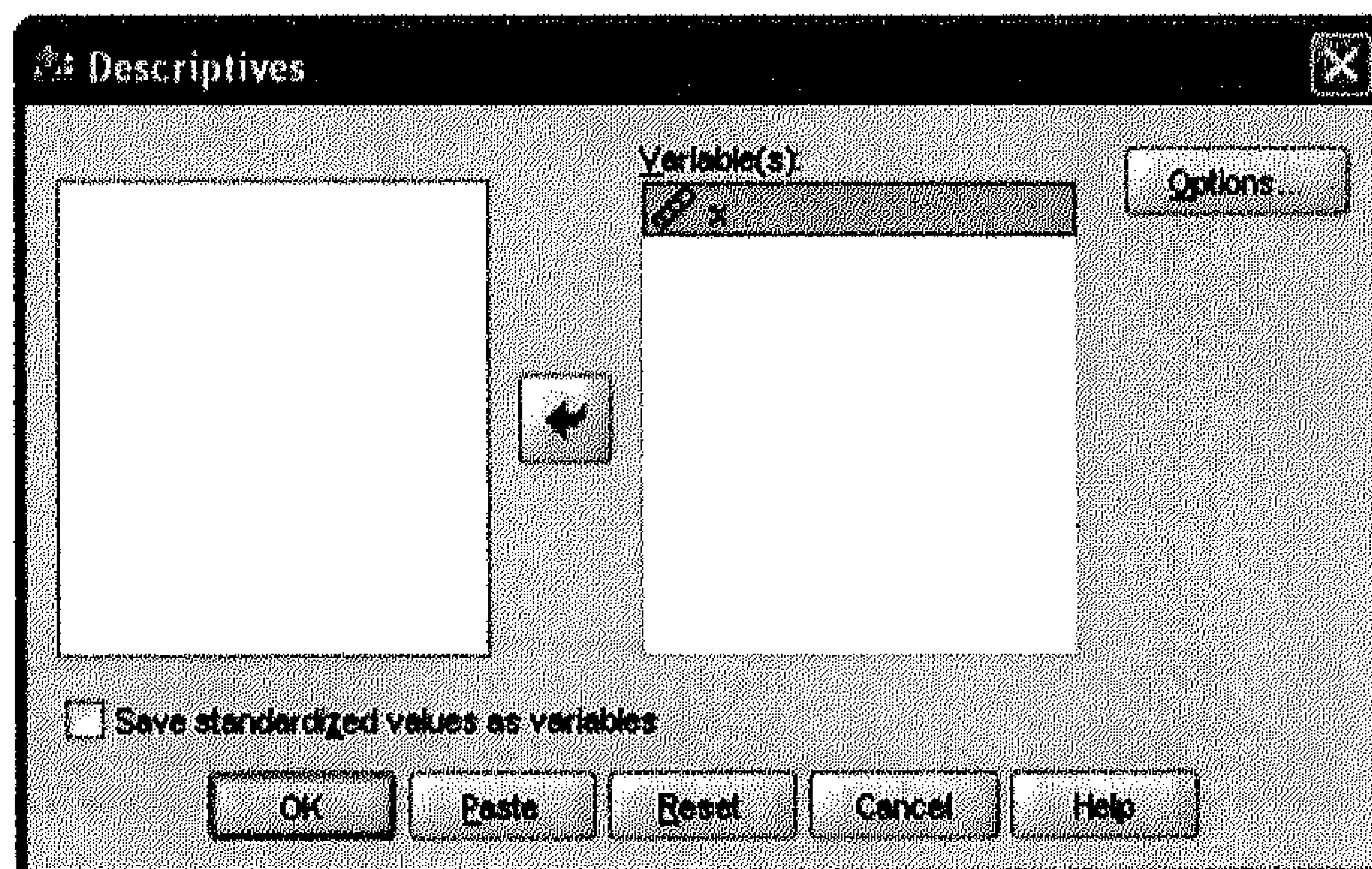
$$Q = \frac{Q_3 - Q_1}{2} = \frac{P_{75} - P_{25}}{2} = \frac{13.5 - 5.5}{2} = \frac{8}{2} = 4$$

(ب) الأمر **Descriptives**

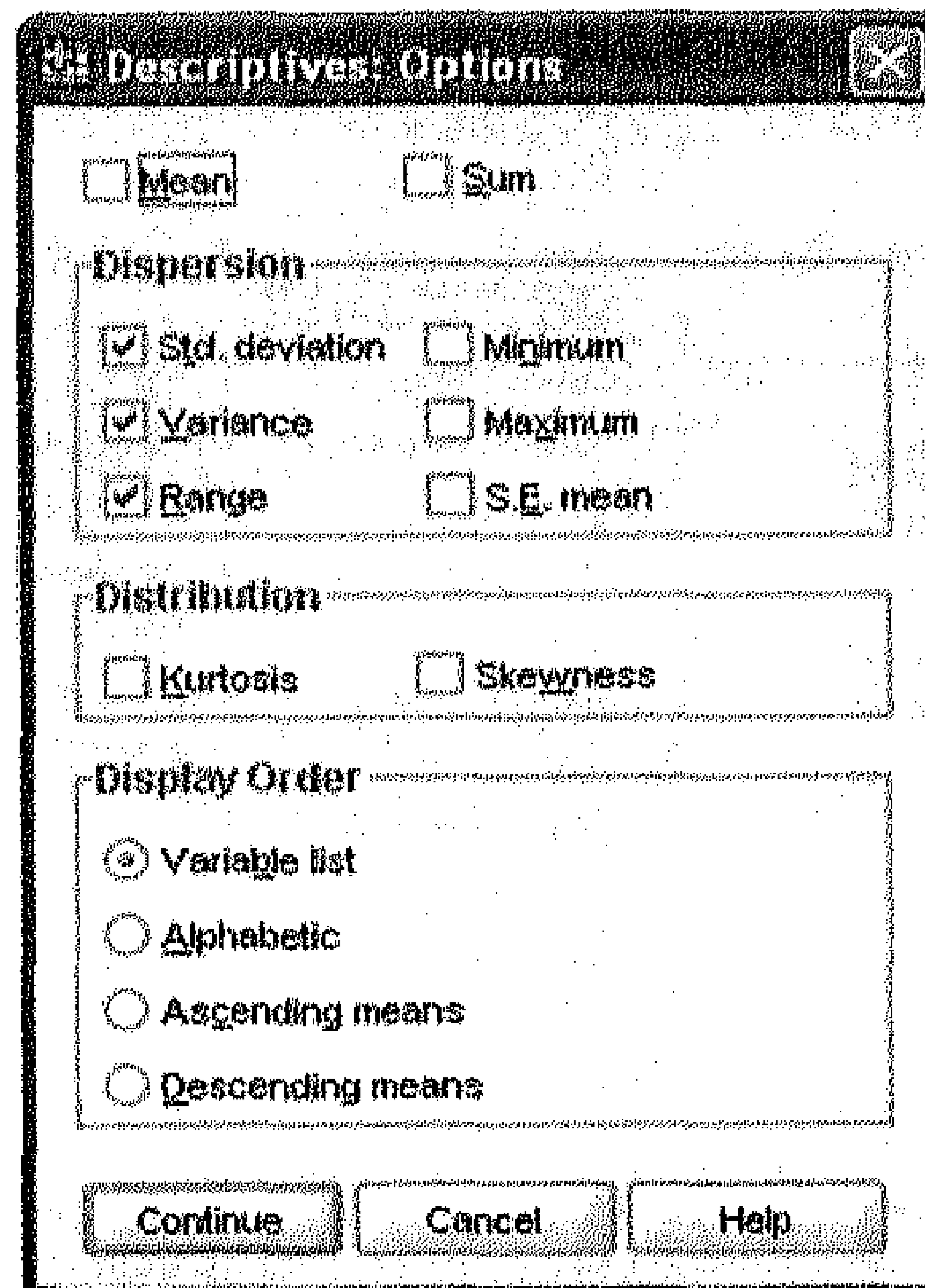
- نضغط على الأمر Descriptives



- تظهر نافذة جديدة بعنوان Descriptives



- ننقل المتغير X لقائمة Variable(s) ثم نضغط على Options فتظهر نافذة جديدة



- من قائمة Dispersion نختار Std. Deviation لحساب الانحراف المعياري، Variance لحساب التباين، Range لحساب المدى،
- ثم نضغط على Continue فنعود للشاشة السابقة، ثم نضغط على Ok فتظهر النتائج التالية:

Descriptive Statistics

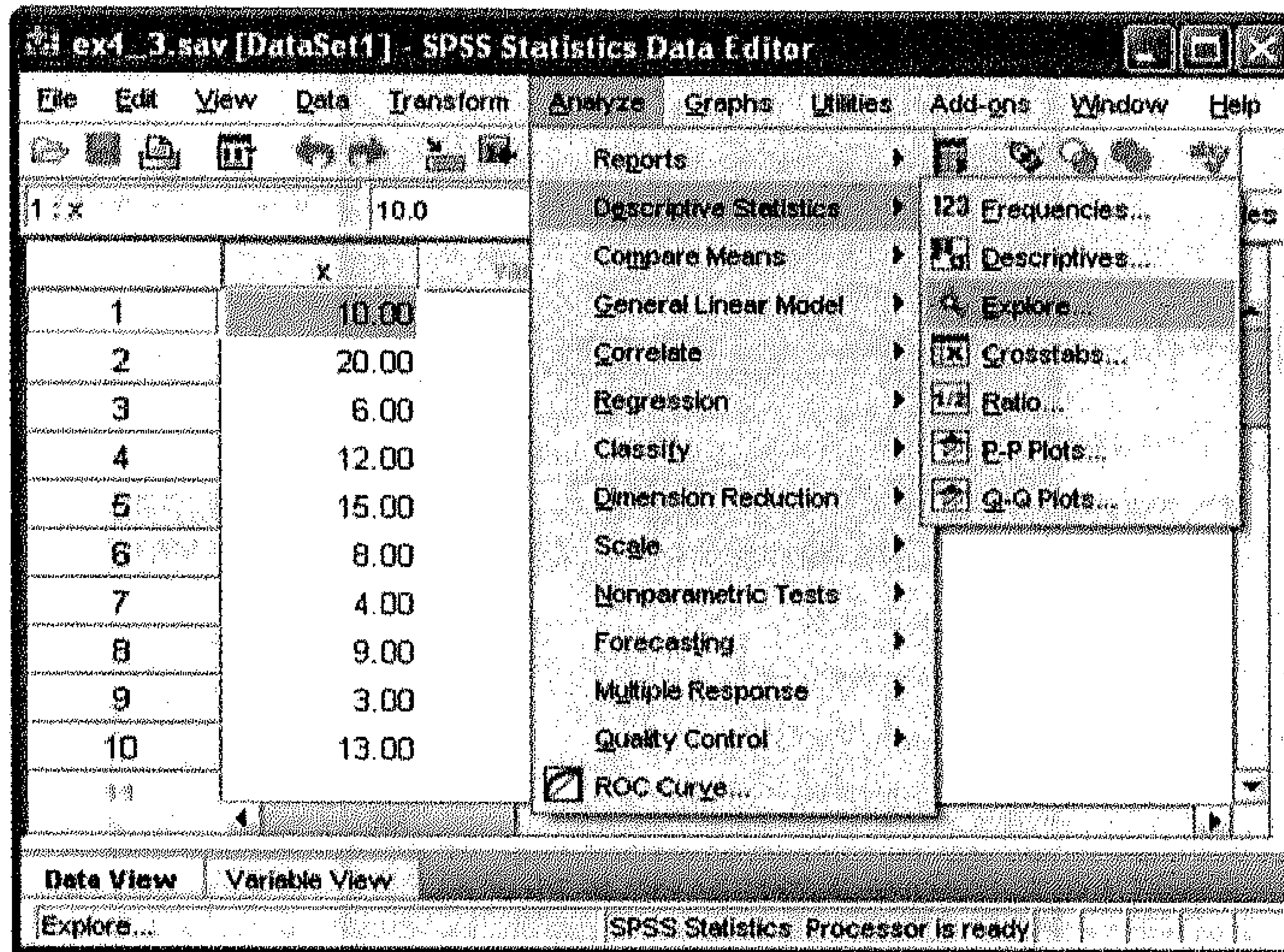
	N	Range	Std. Deviation	Variance
x	10	17.00	5.20683	27.111
Valid N (listwise)	10			

نجد من الجدول السابق أن:

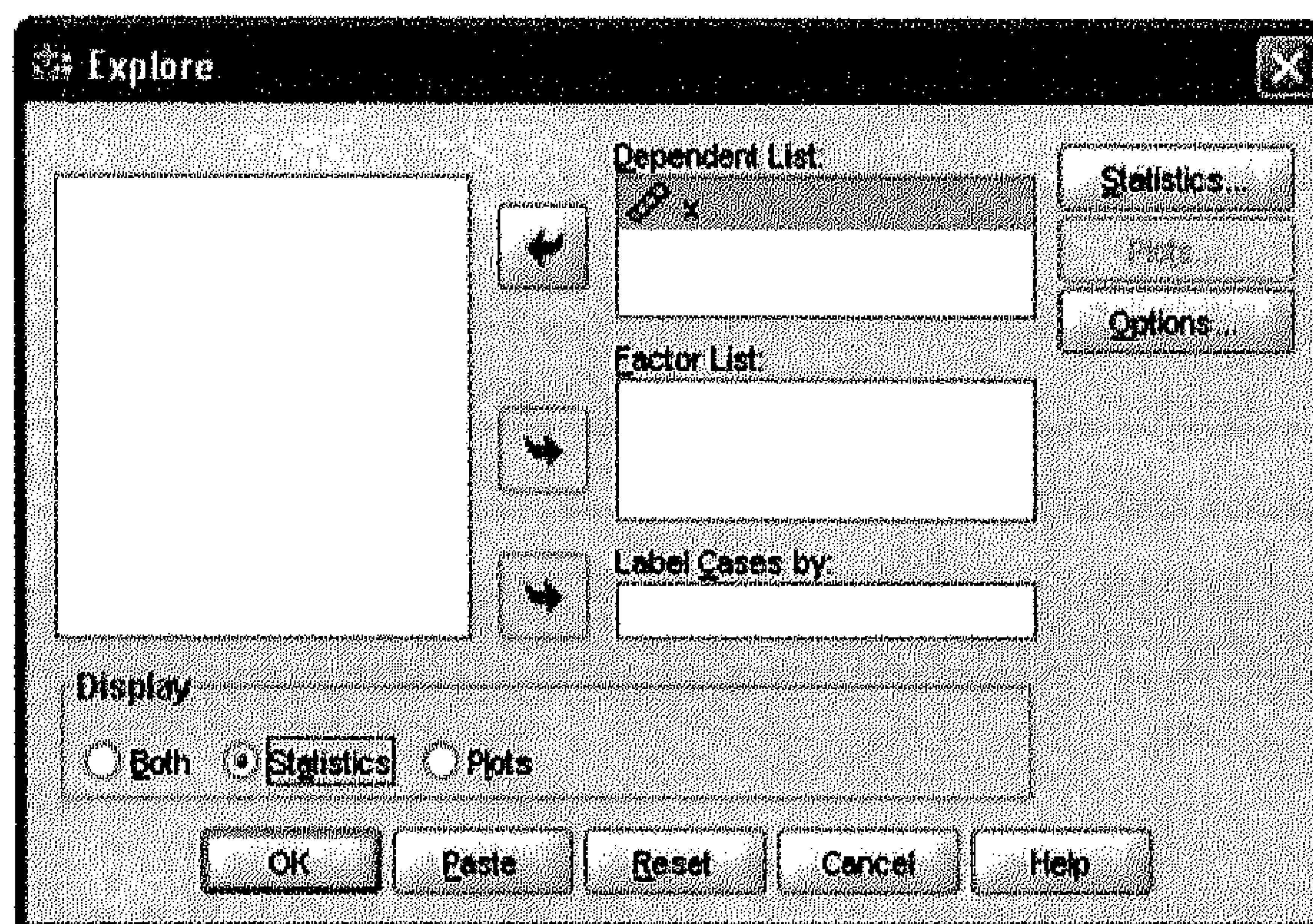
عدد البيانات هو $n = 10$ المدى للبيانات يساوى $Range = 17$، الانحراف المعياري للبيانات هو $s = std.Deviation = 5.20683$ ، التباين للبيانات هو $s^2 = Variance = 27.111$

(ج) الأمر Explore

- نضغط على الأمر Explore

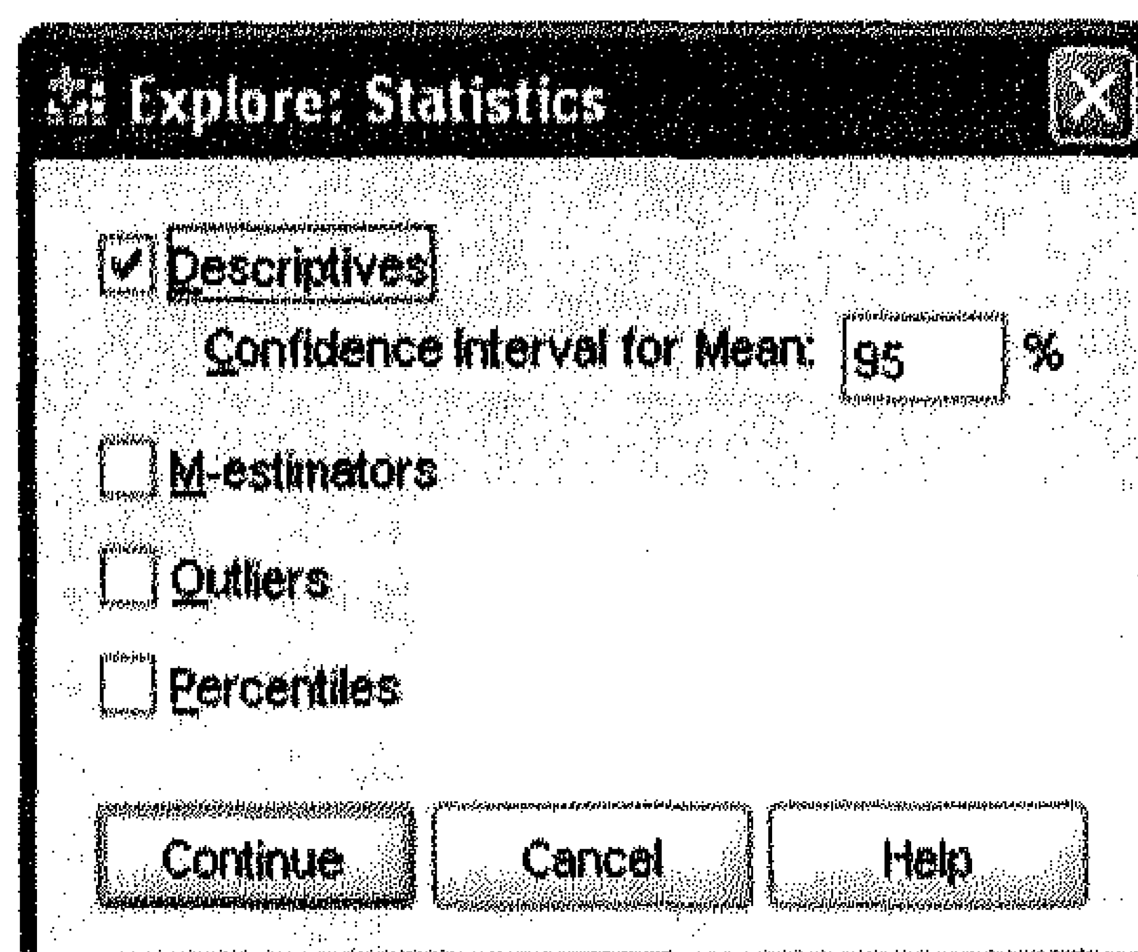


● تظهر نافذة جديدة بعنوان Explore



● ننقل المتغير X لقائمة Dependent List ثم نضغط على Statistics في قائمة Display لحساب الاحصاءات فقط.

● نضغط على Statistics أعلى يمين النافذة لتحديد الاحصاءات المراد حسابها فتظهر نافذة جديدة:



- نختار الأمر Descriptives ثم نضغط على Continue فنعود للشاشة السابقة ثم نضغط على Ok فتظهر النتائج التالية:

Descriptives

			Statistic	Std. Error
x	Mean		10.0000	1.64655
	95% Confidence Interval for Mean	Lower Bound	6.2753	
		Upper Bound	13.7247	
	5% Trimmed Mean		9.8333	
	Median		9.5000	
	Variance		27.111	
	Std. Deviation		5.20683	
	Minimum		3.00	
	Maximum		20.00	
	Range		17.00	
	Interquartile Range		8.00	
	Skewness		.519	.687
	Kurtosis		.023	1.334

نجد من الجدول السابق أن الوسط الحسابي $\bar{x} = Mean = 10$، الوسيط $Med = 9.5$ ، التباين للبيانات $s^2 = Variance = 27.111$، الانحراف المعيارى $s = Std.Deviation = 5.20683$ ، أقل قيمة في البيانات وأكبر قيمة هي $X_{max} = Maximum = 20$ ، $X_{min} = Minimum = 3.00,$ ، المدى للبيانات

يساوى $Range = 17$، المدى الربيعى للبيانات هو $Q_3 - Q_1 = interquartile\ Range = 8.00,$ ، معامل الالتواء هو $sk = skewness = 0.519$ ونلاحظ أنه موجب وبالتالي فإن المنحنى ملتوٍ ناحية اليمين، معامل التفرطح مقارنة بالتوزيع الطبيعي هو$k - 3 = kurtosis = 0.023$ ونجد أن هذا الفرق موجب لذا فإن المنحنى مدبب.

تطبيق (٤-٣):

من مثال (٤-٤): لعينة من 500 أسرة اختيرت من إحدى المدن توفرت لدينا البيانات التالية عن عدد أفراد الأسرة

عدد أفراد الأسرة	2	3	4	5	6	7	8	9
عدد الأسر	98	118	101	95	50	20	10	8

احسب

(أ) التباين والانحراف المعياري لعدد أفراد الأسرة.

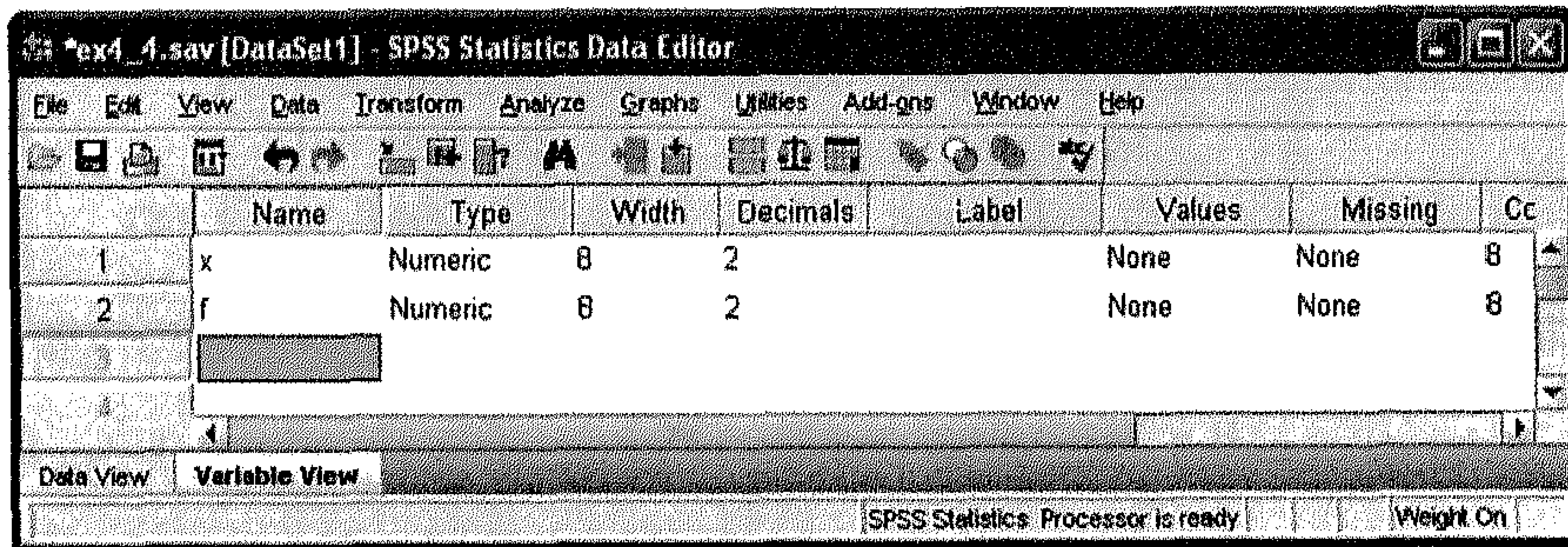
(ب) معامل الالتواء والتفرطح.

(ج) الدرجات المعيارية لعدد أفراد الأسرة.

الحل

أولاً: سوف نقوم بإدخال البيانات للبرنامج باتباع الخطوات التالية:

١- من نافذة Variable View نعرف متغيرين x, f



٢- ثم ننتقل لنافذة Data View ونقوم بإدخال البيانات.

*ex4_4.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

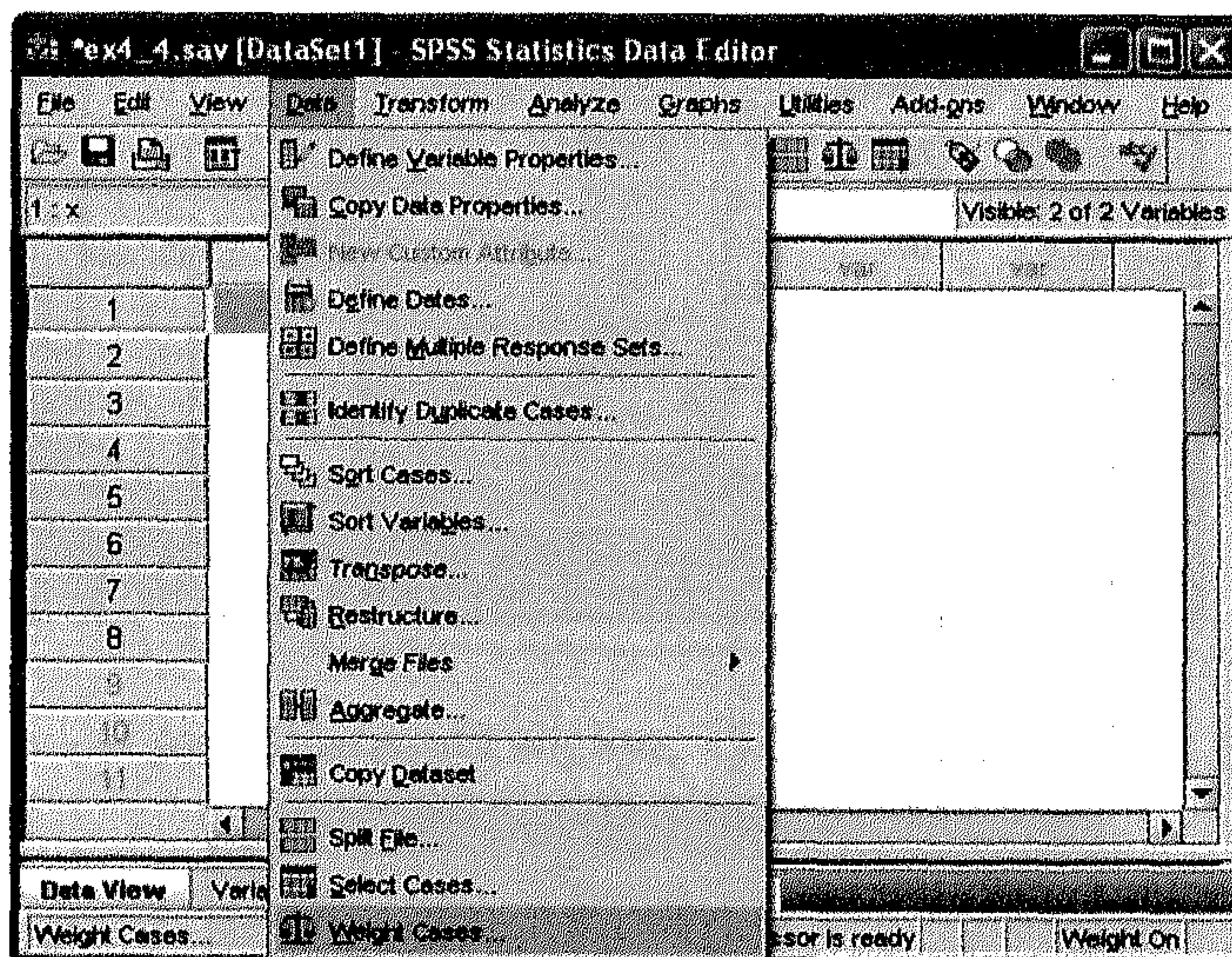
1 : x 2.0 Visible: 2 of 2 Variables

	x	f
1	2.00	98.00
2	3.00	118.00
3	4.00	101.00
4	5.00	95.00
5	6.00	50.00
6	7.00	20.00
7	8.00	10.00
8	9.00	8.00

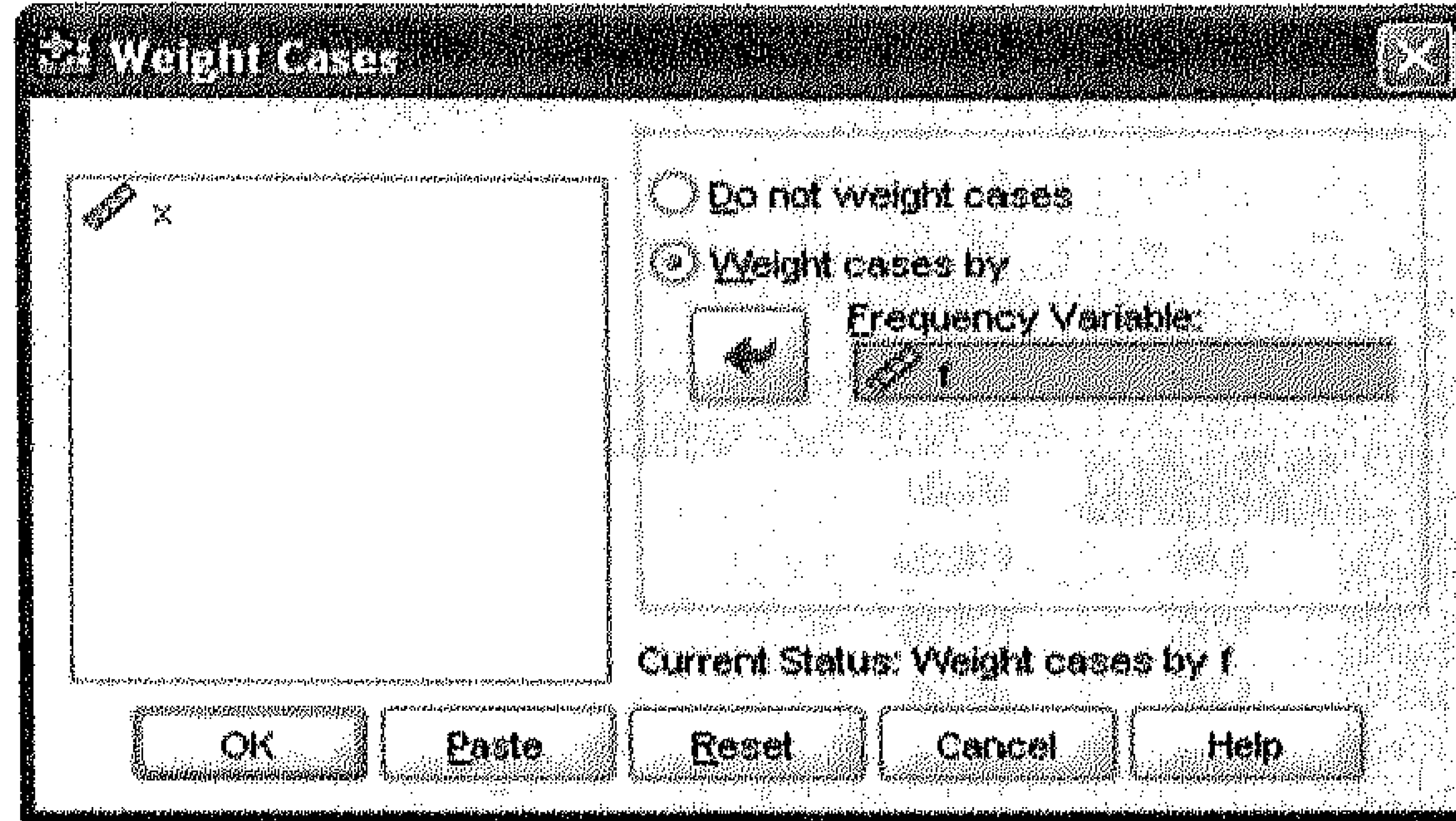
Data View Variable View SPSS Statistics Processor is ready Weight On

ثانياً: نحدد أن f هي تكرار للمتغير x وذلك باتباع الخطوات التالية:

١- من قائمة Data نختار Weight Cases



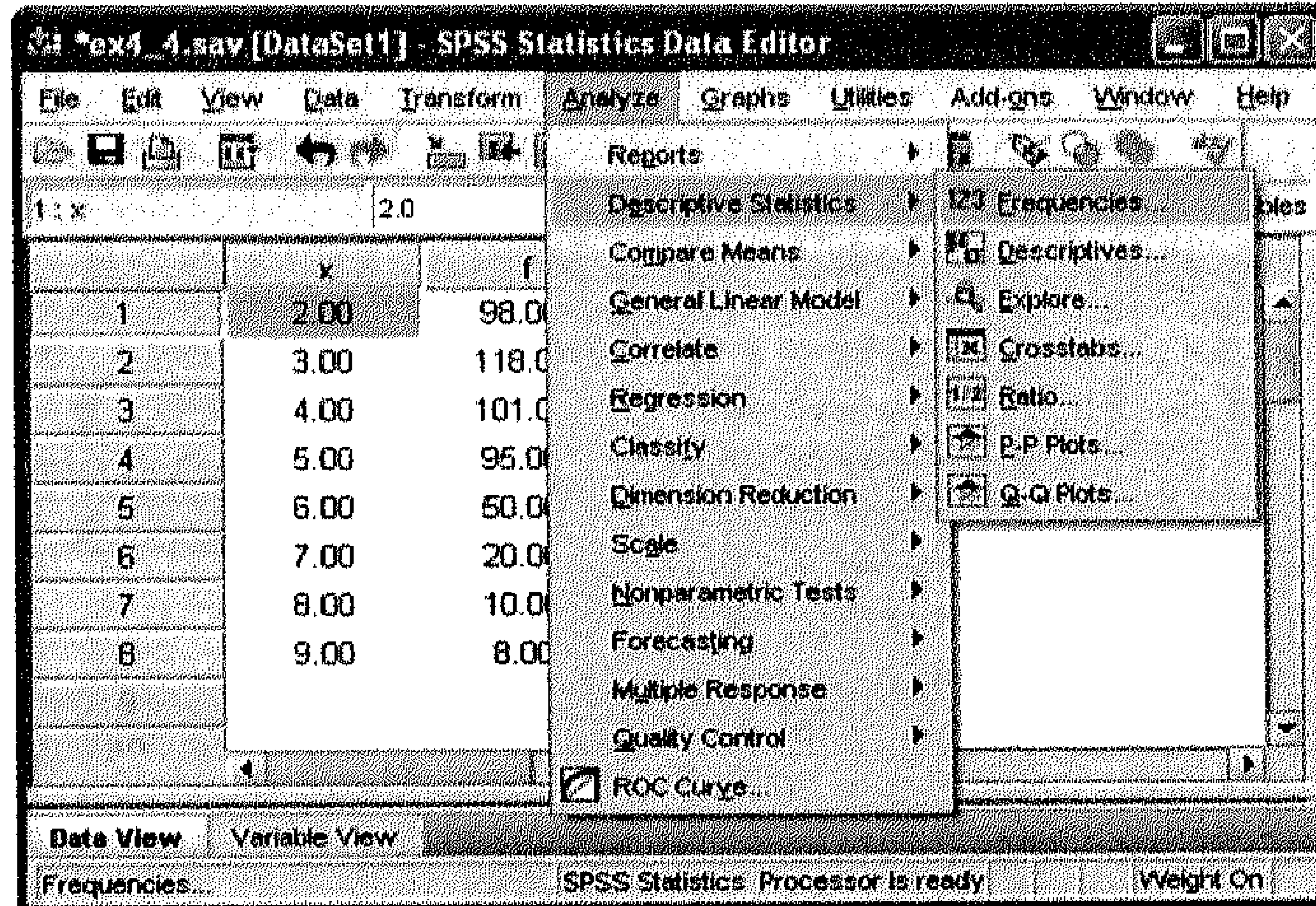
٢- نختار من نافذة Weight Cases الاختيار Weight cases by ثم ننقل المتغير f لخانة Frequency Variable



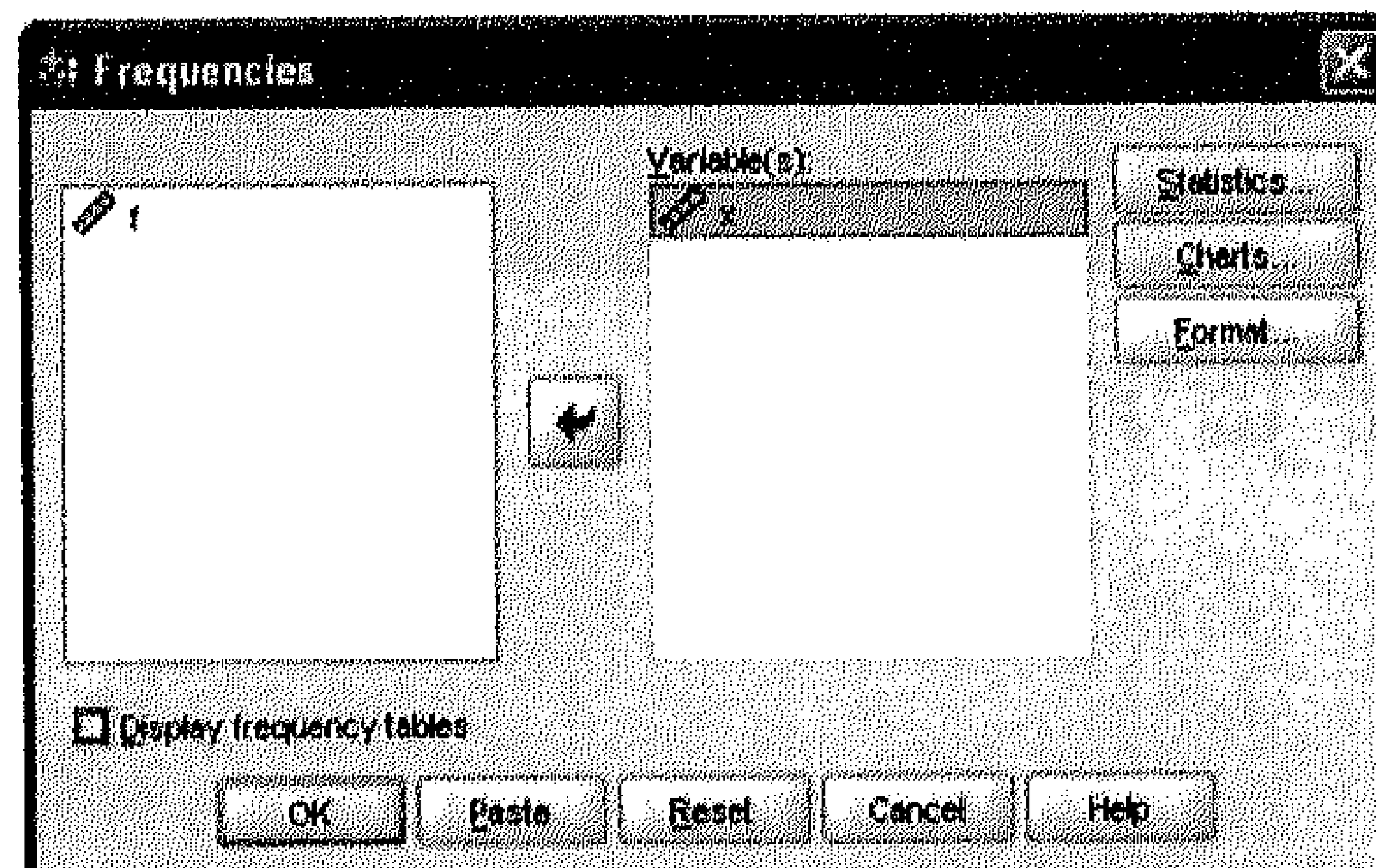
٣- نضغط على ok فنعود لملف البيانات

ثالثاً: تعيين كلاً من التباين والانحراف المعياري ومعامل الالتواء والتفرطح وذلك باتباع الخطوات التالية:

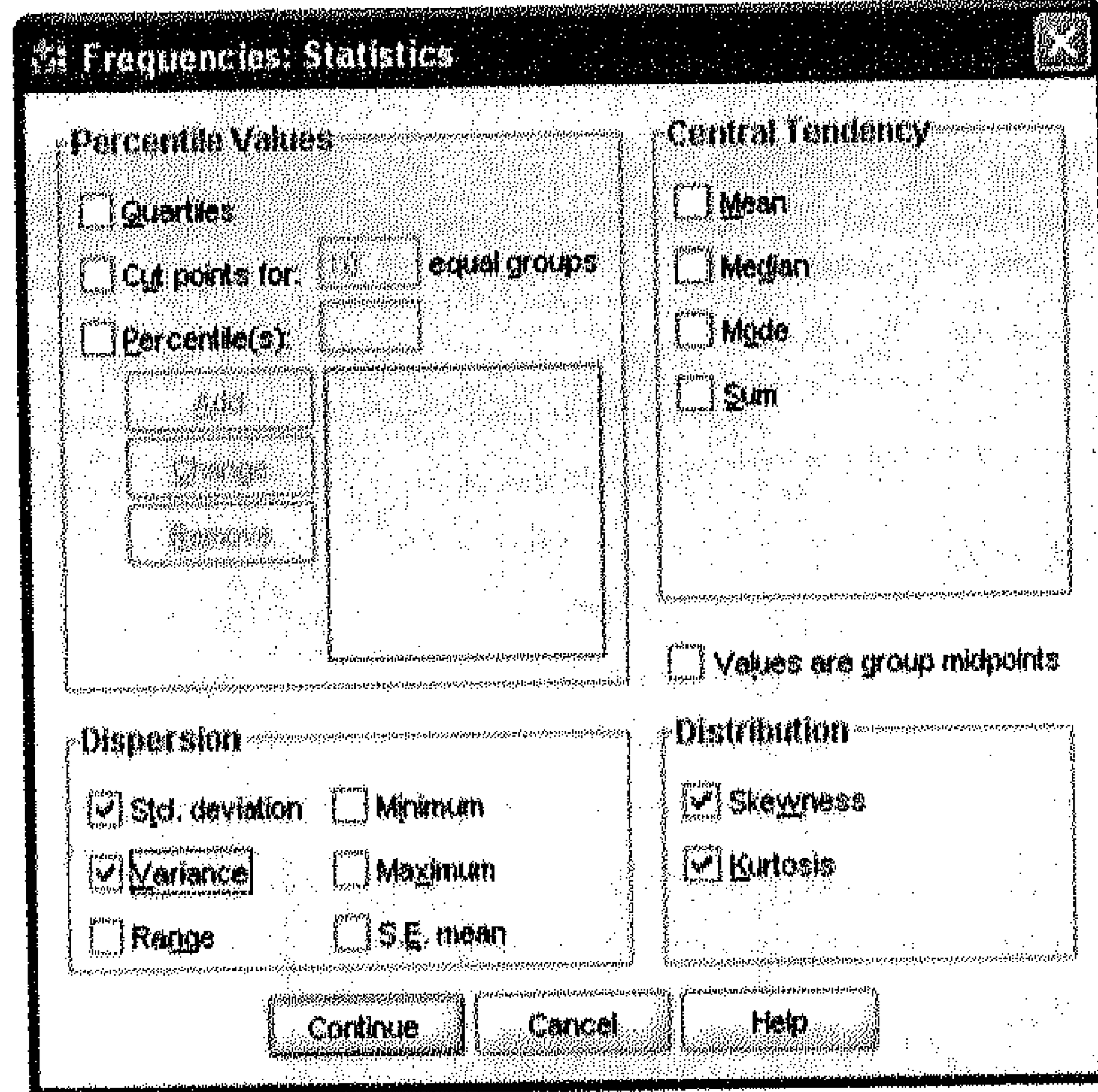
١- من قائمة Analyze نختار Descriptive Statistics فتظهر قائمة منسدلة نختار منها Frequencies



٢- تظهر نافذة جديدة بعنوان Frequencies ننقل المتغير x لقائمة Variable(s) ثم نختار Statistics



٣- من نافذة Frequencies: Statistics سوف نختار كلاً من St. Deviation, Variance من قائمة Dispersion وأيضا نختار Skewness, Kurtosis من قائمة Distribution



٤- نضغط على Continue فنعود للشاشة السابقة ثم نضغط Ok فتظهر النتائج التالية:

Statistics

x

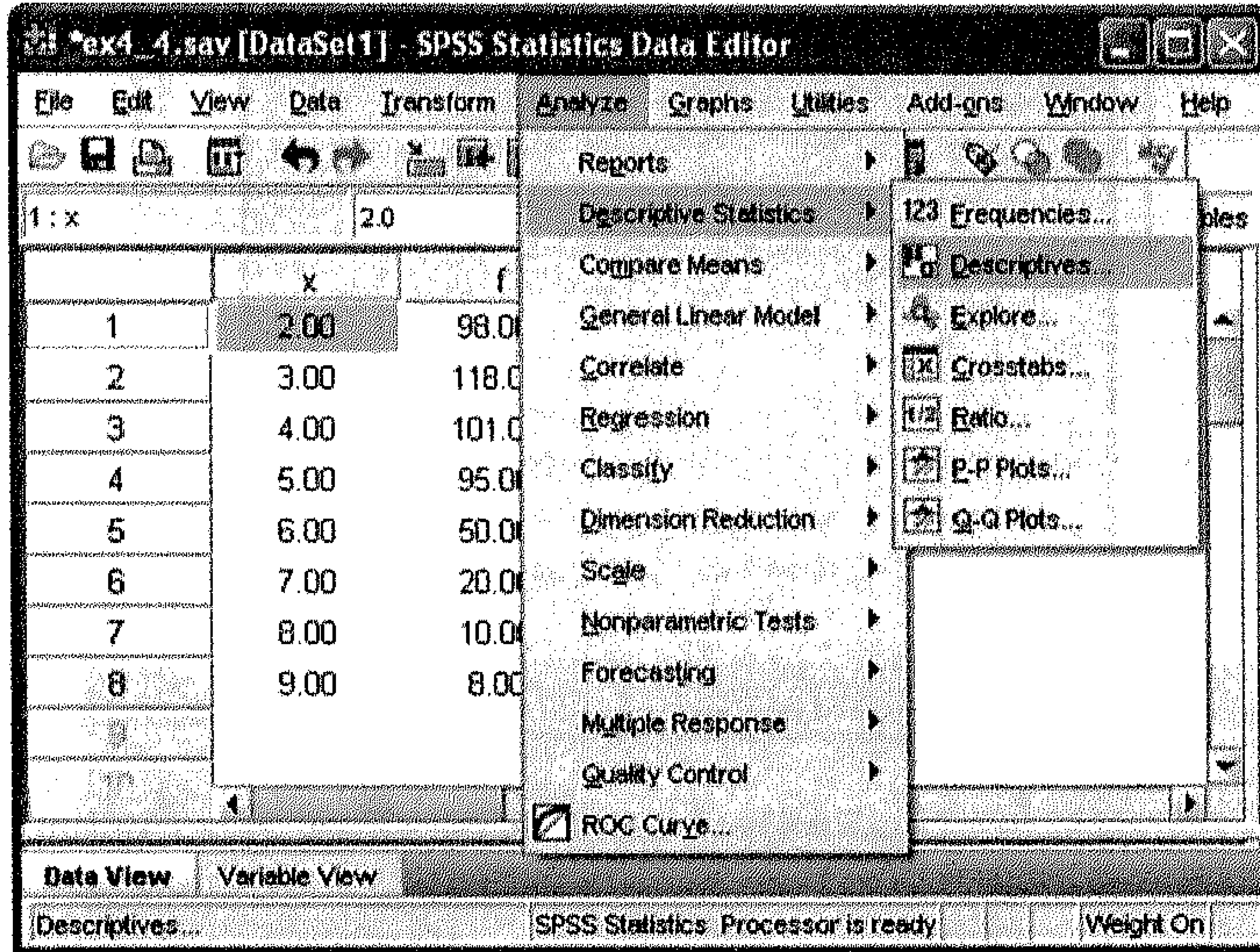
N	Valid	500
	Missing	0
Std. Deviation		1.64123
Variance		2.694
Skewness		.730
Std. Error of Skewness		.109
Kurtosis		.190
Std. Error of Kurtosis		.218

الجدول الأول يحتوى على عدد البيانات والذي يساوى مجموع التكرارات $n = 500$ الانحراف المعياري $s = 1.64123$ ، التباين $s^2 = 2.694$ ، معامل الالتواء $sk = 0.73$ ونلاحظ أن الالتواء موجب وبالتالي فإن المنحنى ملتو ناحية اليمين، ومعامل التفرطح مقارنة بالتوزيع الطبيعي $k - 3 = kurtosis = 0.190$ ونجد أن هذا الفرق موجب وبالتالي فإن منحنى البيانات مدبب.

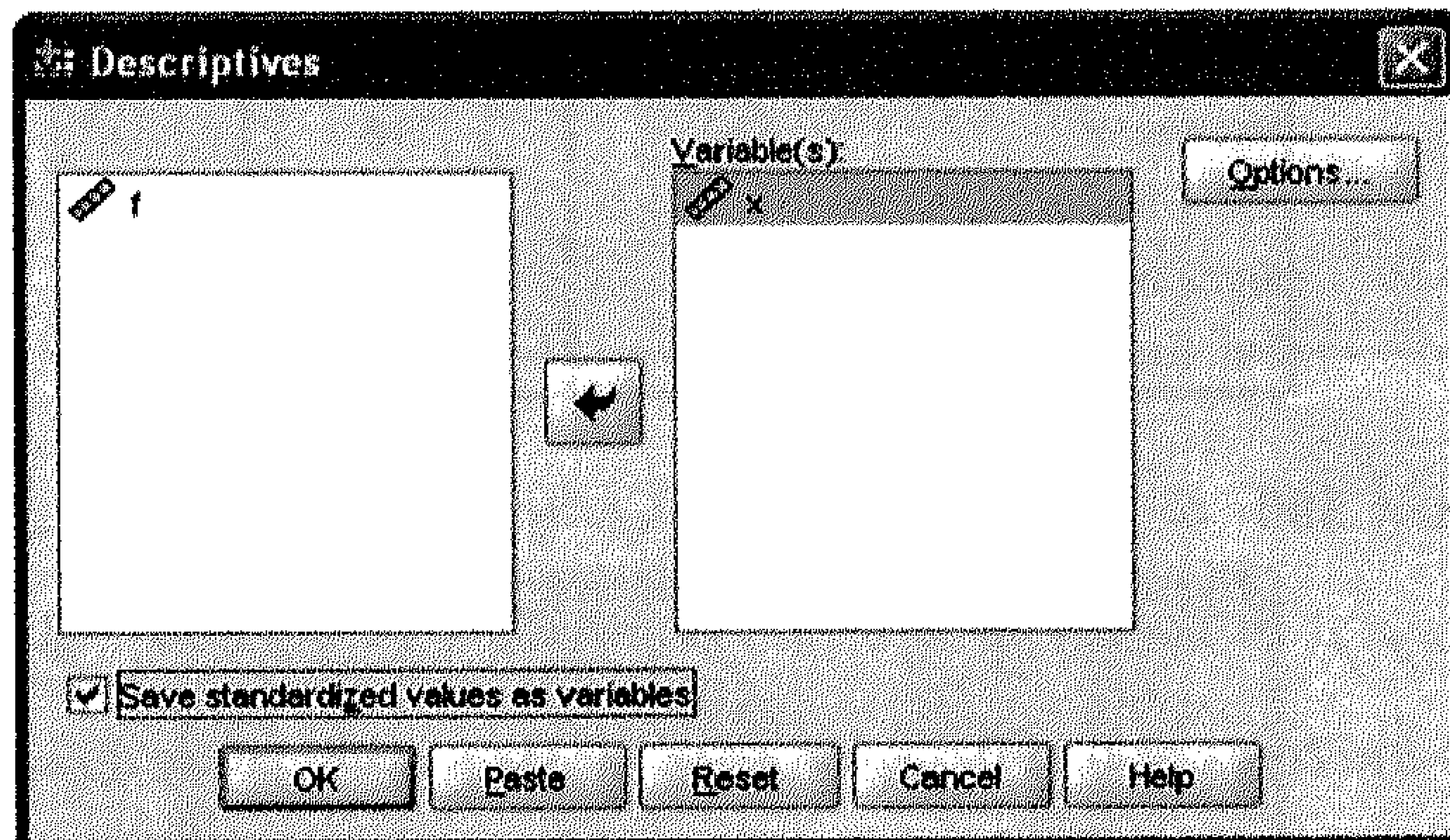
رابعاً: لتعيين الدرجات المعيارية لعدد أفراد الأسرة نتبع الخطوات التالية:

١- من قائمة Analyze نختار Descriptive Statistics

٢- ومن القائمة المنسدلة نختار Descriptives

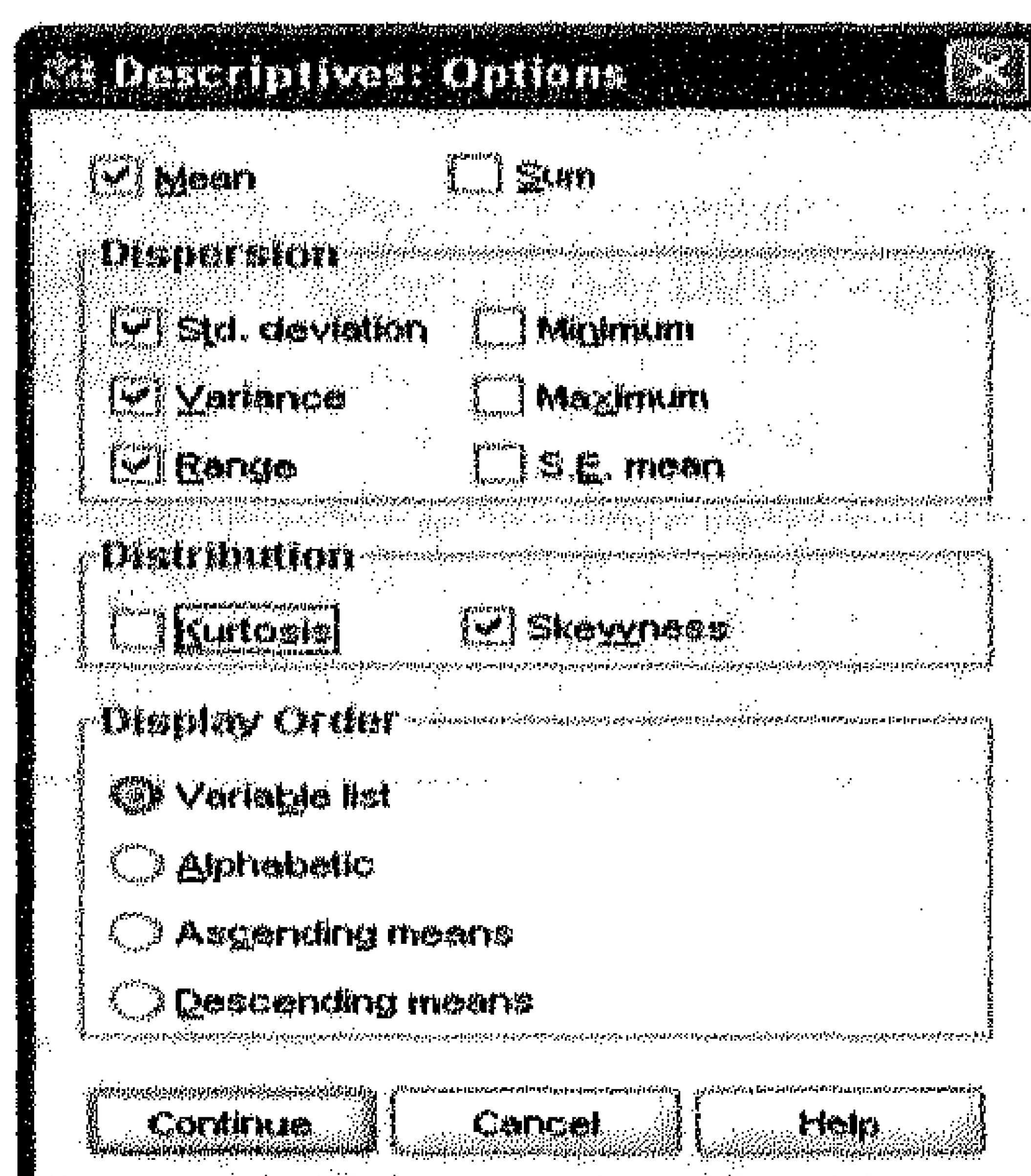


٣- تظهر نافذة جديدة بعنوان Frequencies ننقل المتغير x لقائمة Variable(s) ثم نختار Save standardized values as variables لتعيين الدرجات المعيارية



٤- ثم نضغط على Options فتظهر شاشة جديدة بعنوان Descriptives Options

٥- نختار منها المقاييس التي نرغب في تعيينها ثم نختار Continue



٦- فنعود للشاشة السابقة نضغط Ok فتظهر النتائج التالية:

Descriptive Statistics

	N	Range	Mean	Std. Deviation	Variance
x	500	7.00	4.0420	1.64123	2.694
Valid N (listwise)	500				

والجدول السابق يحتوى على المقاييس التى تم تحديدها من نافذة Options وهى عدد القيم $n = 500$ أقل قيمة فى البيانات وأكبر قيمة $X_{min} = 2, X_{max} = 9$ الوسط الحسابى $\bar{x} = Mean = 4.042$ الانحراف المعيارى $s = std.Devation = 1.64123$ وبالعودة لملف البيانات

*ex4_4.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : Zx -1.24419173492551 47 Visible: 3 of 3 Variables

	x	f	Zx
1	2.00	98.00	-1.24419
2	3.00	118.00	-0.63489
3	4.00	101.00	-0.02559
4	5.00	95.00	0.58371
5	6.00	50.00	1.19301
6	7.00	20.00	1.80231
7	8.00	10.00	2.41161
8	9.00	8.00	3.02091

Data View Variable View

SPSS Statistics Processor is ready Weight On

نلاحظ أن البرنامج قد أضاف عموداً جديداً يسمى z_x وهى الدرجات المعيارية للمتغير x فنجد أن القيمة $x_1 = 2$ يقابلها درجة $z_1 = -1.24419$ وتلك القيمة لها التكرار $f_1 = 98$ وهكذا لباقي القيم(البيانات).

تطبيق (٤-٤)

من مثال (٤-٥): الجدول التكراري الآتي يبين توزيع 100 عمود من الصلب حسب أطوالها (إلى أقرب 0.01 بوصة) احسب

فئات الطول	3.8–	3.9–	4.0–	4.1–	4.2–	4.3–	4.4–
التكرار	3	8	14	19	28	18	10

(أ) المدى.

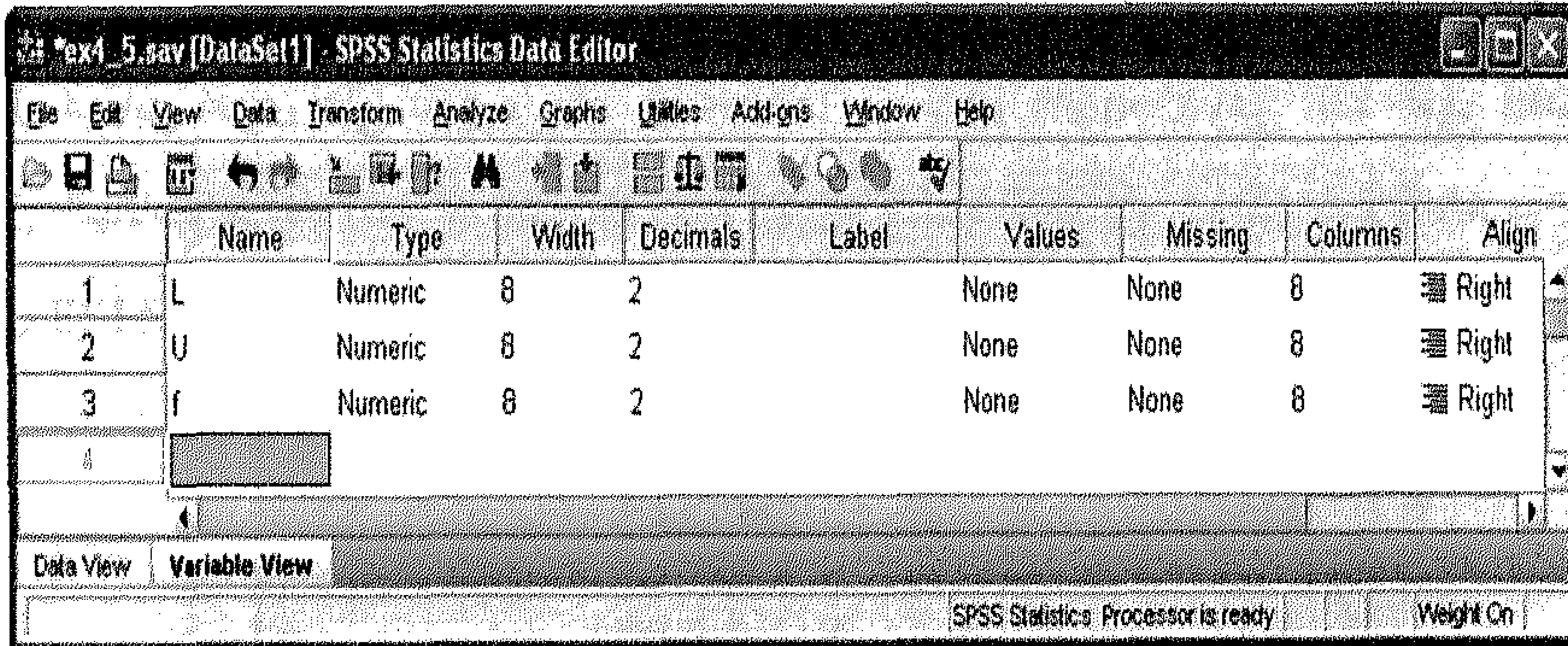
(ب) التباين والانحراف المعياري لطول العمود.

(ج) معامل الاختلاف لطول العمود.

(د) معامل الالتواء والتفرطح.

الحل

١- نقوم بتعريف ثلاثة متغيرات هم U, L, f تمثل حدود الفترات والتكرار المقابل لكل فترة


*ex4_5.sav [DataSet1] - SPSS Statistics Data Editor

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align
1	L	Numeric	8	2		None	None	8	Right
2	U	Numeric	8	2		None	None	8	Right
3	f	Numeric	8	2		None	None	8	Right
4									

٢- ثم نقوم بإدخال حدود الفترات الحد الأدنى للفترة L والحد الأعلى للفترة U، تكرار الفترات f

*ex4_5.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : L 3.8 Visible: 3 of 3 Variables

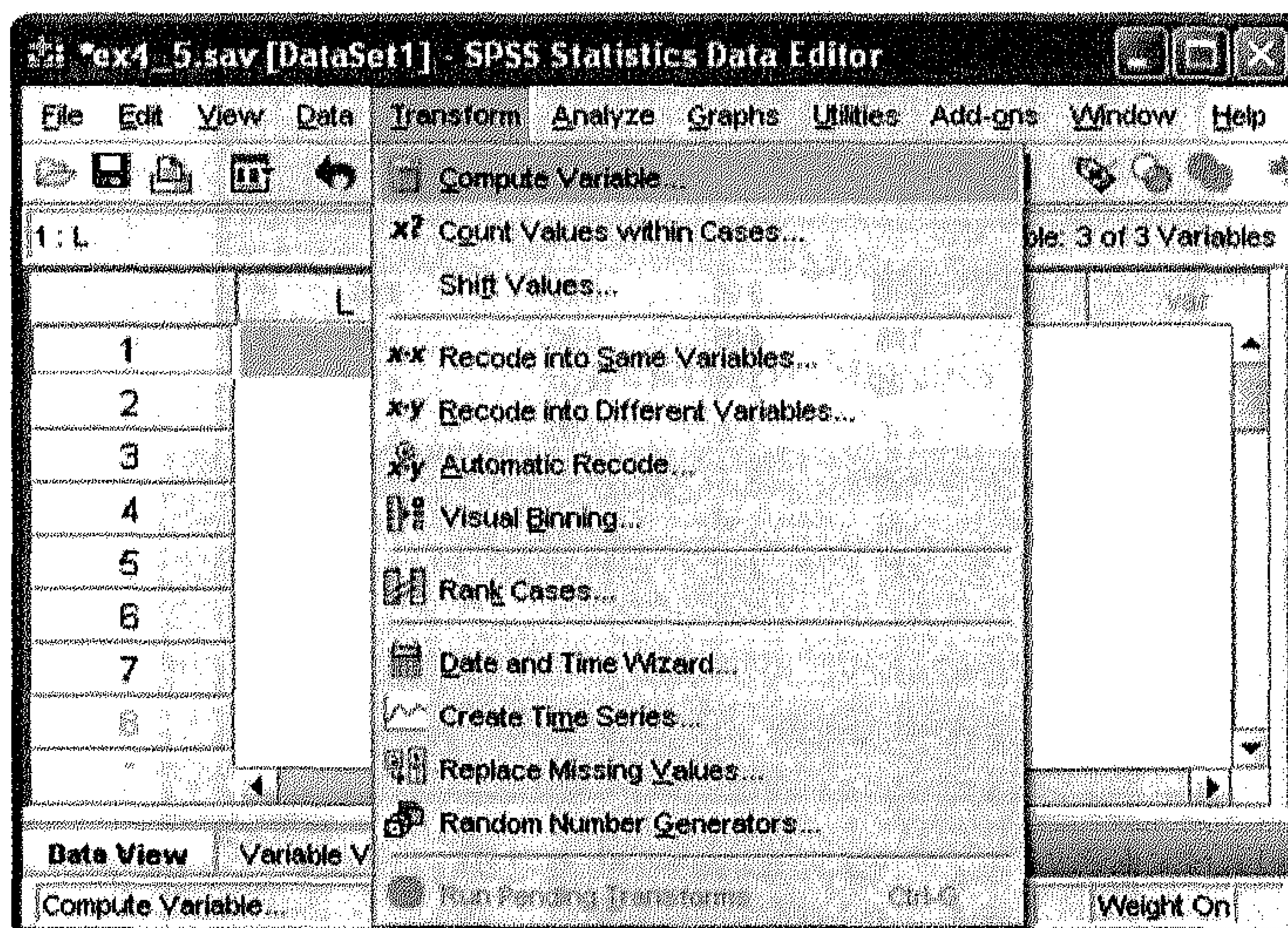
	L	U	f	var	var
1	3.80	3.90	3.00		
2	3.90	4.00	8.00		
3	4.00	4.10	14.00		
4	4.10	4.20	19.00		
5	4.20	4.30	28.00		
6	4.30	4.40	18.00		
7	4.40	4.50	10.00		
8					

Data View | Variable View

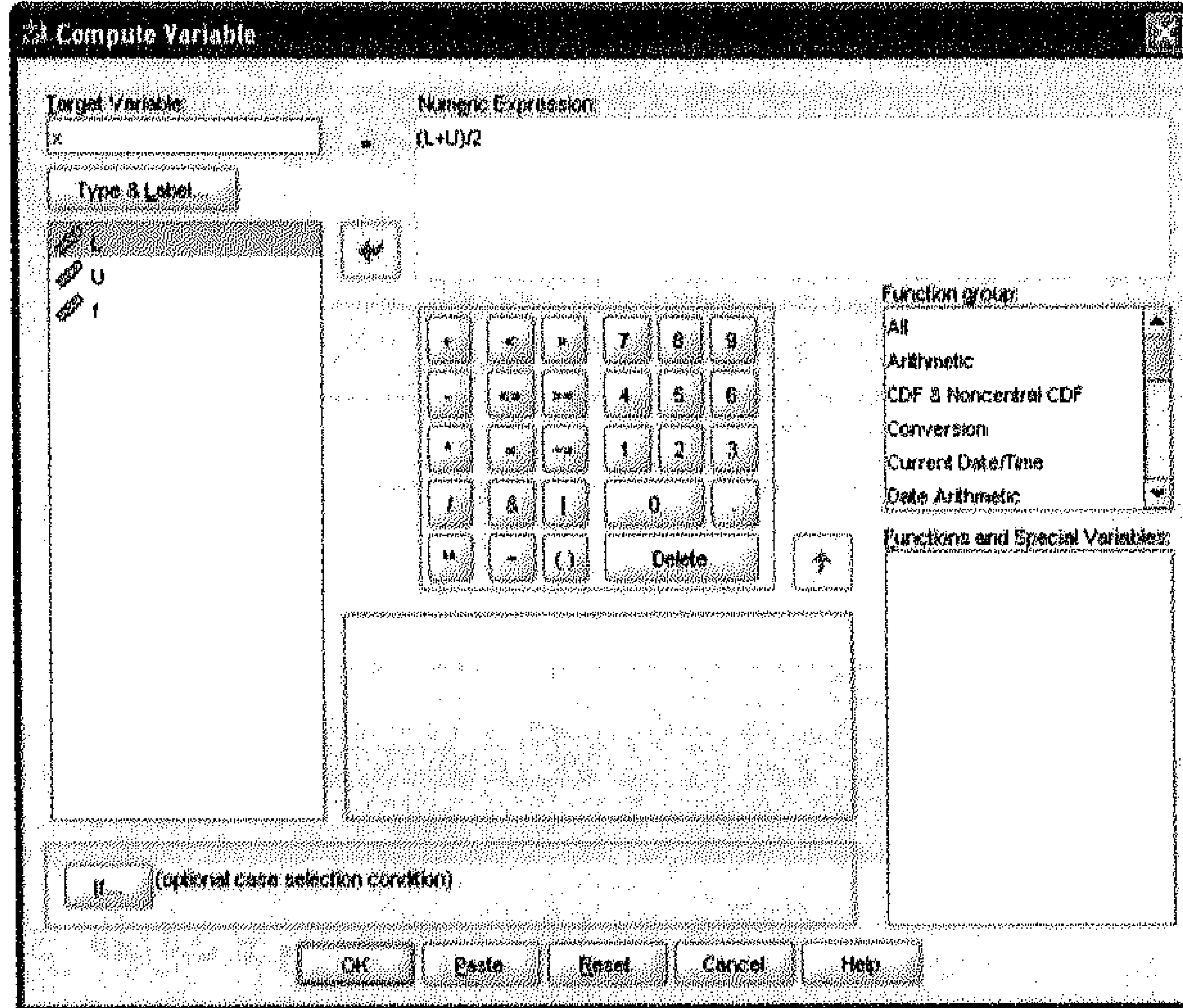
SPSS Statistics Processor is ready | Weight On

٣- يجب حساب مركز الفترات x وذلك باتباع الخطوات التالية:

- من قائمة Transform نختار الأمر Compute Variable



- سوف تظهر شاشة جديدة بعنوان Compute Variable نكتب اسم المتغير الجديد x في خانة Target Variable
- نكتب العلاقة التي يحسب منها مركز الفترات $\frac{L+U}{2}$ في خانة Numeric Expression ثم نختار Ok



- سوف نعود لملف البيانات وقد أضاف متغير جديد x يحتوى على مركز الفترات

*ex4_5.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : L 3.8 Visible: 4 of 4 Variables

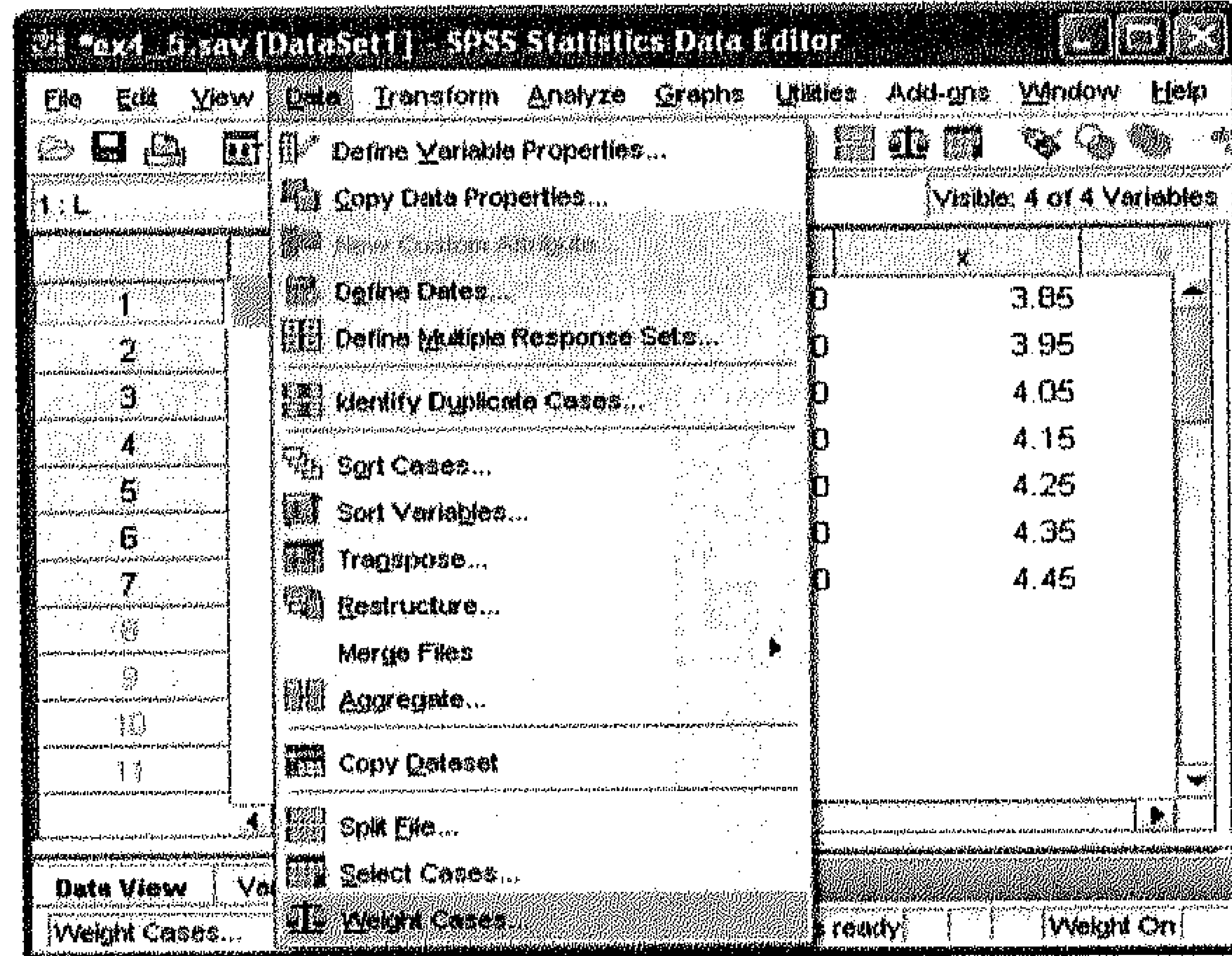
	L	U	f	x
1	3.80	3.90	3.00	3.85
2	3.90	4.00	8.00	3.95
3	4.00	4.10	14.00	4.05
4	4.10	4.20	19.00	4.15
5	4.20	4.30	28.00	4.25
6	4.30	4.40	18.00	4.35
7	4.40	4.50	10.00	4.45
8				

Data View | Variable View

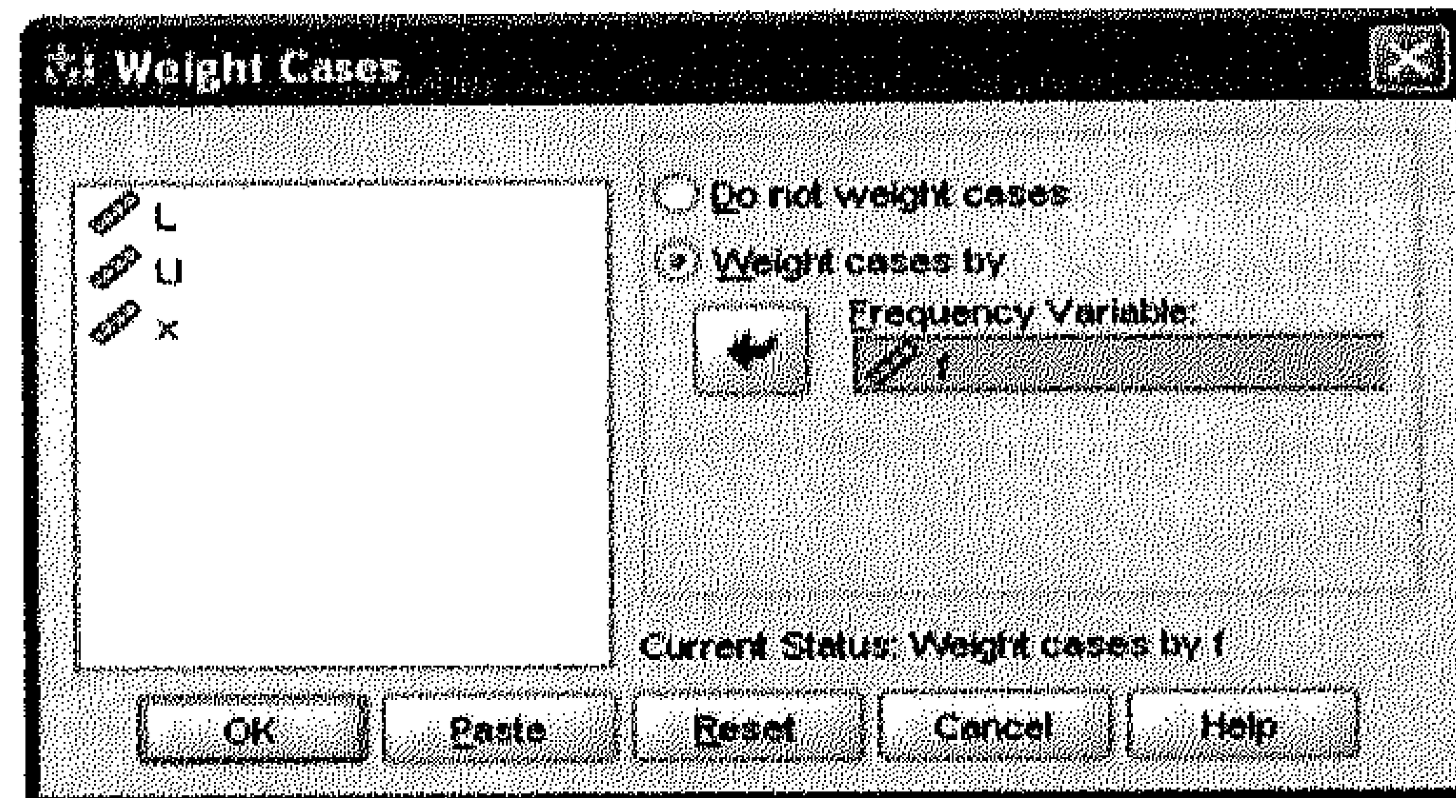
SPSS Statistics Processor is ready | Weight On

٤- يجب تعريف أن f هي تكرار الفترات وذلك بإتباع الخطوات التالية:

- باستخدام الأمر Weight Cases من قائمة Data

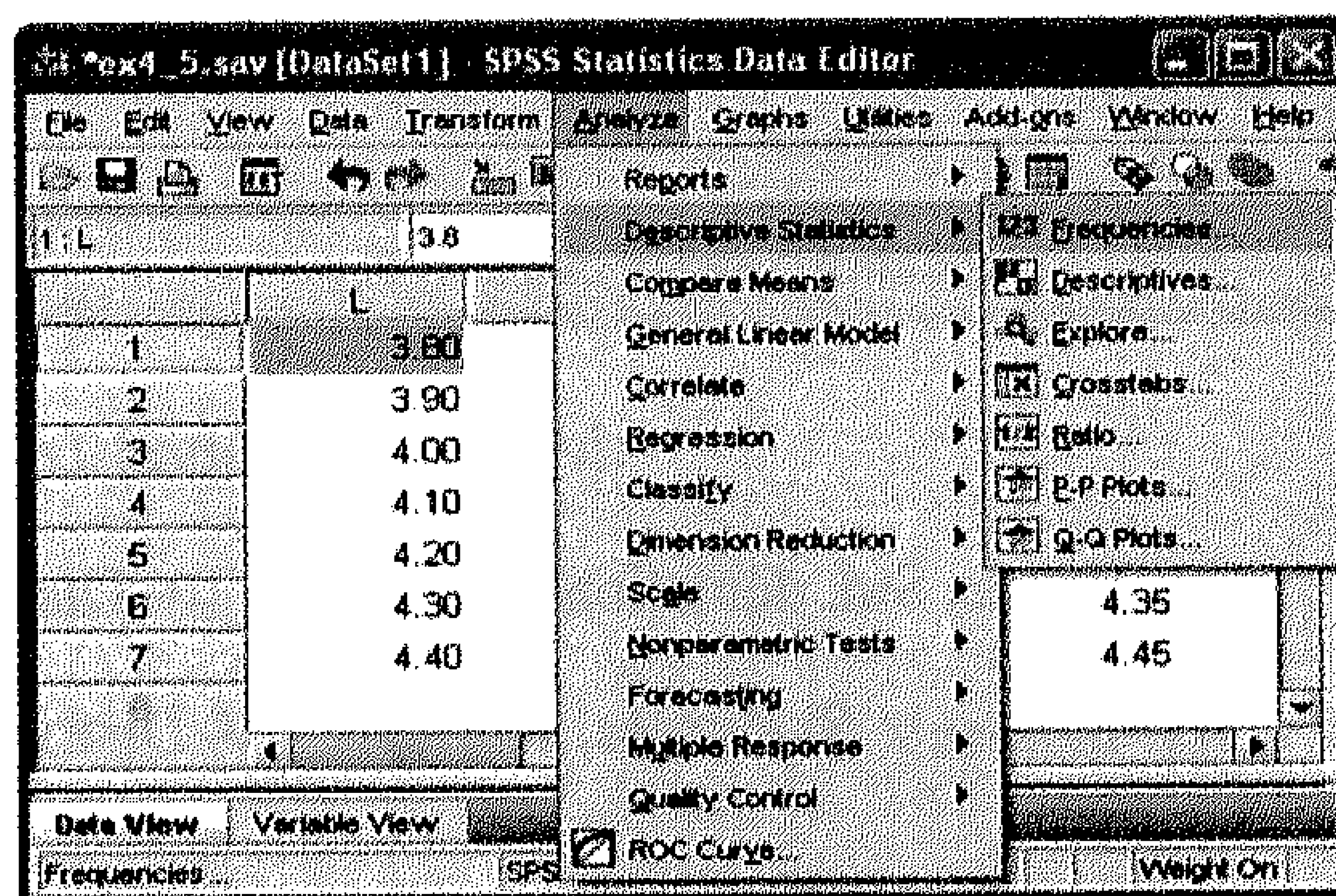


- من نافذة Weight Cases نختار Weigh Cases by ثم ننقل المتغير f لخانة Frequency Variable ثم نضغط على Ok فنعود لملف البيانات

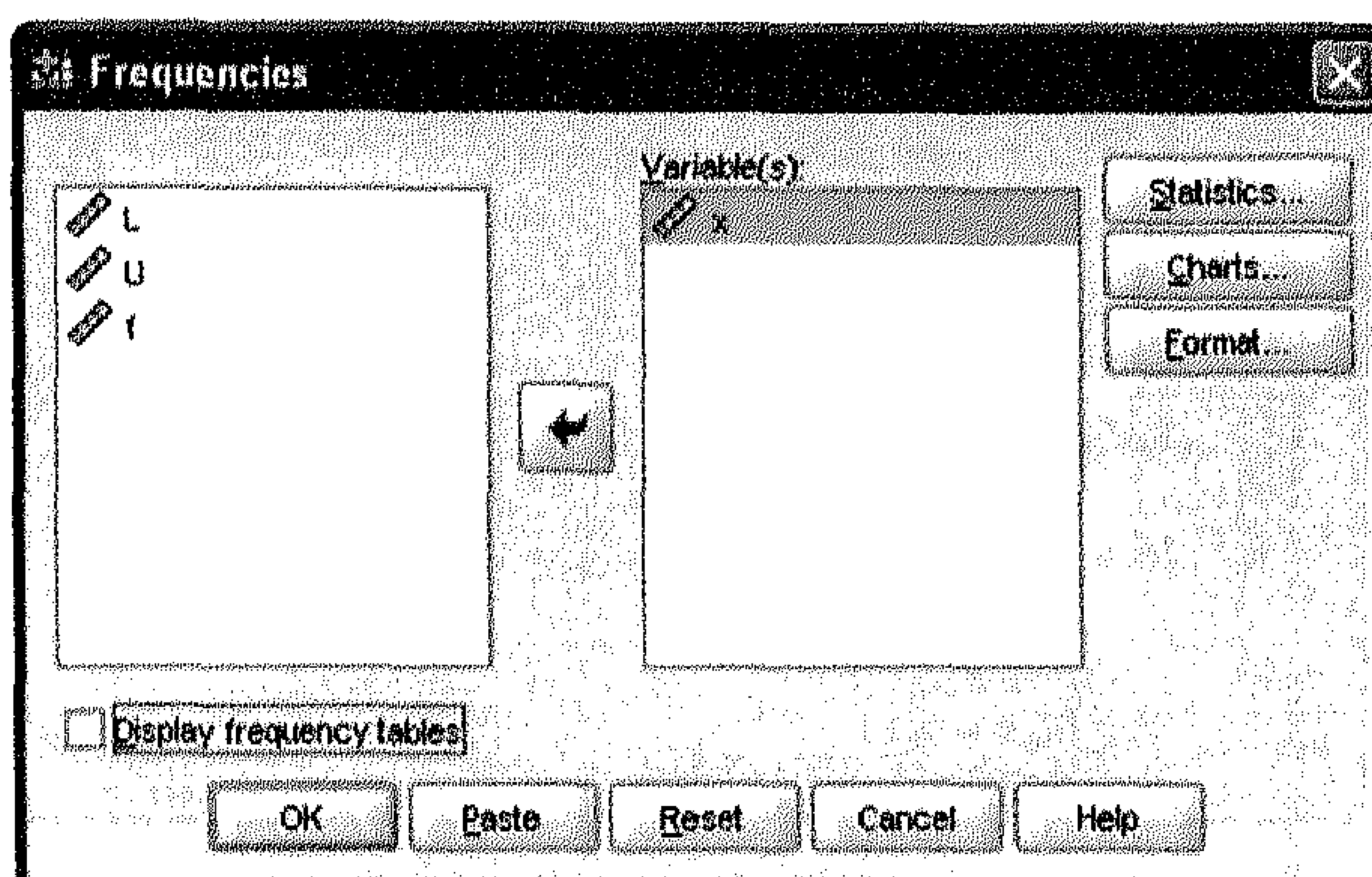


٥- لتعيين المقاييس المطلوبة نتبع الخطوات التالية:

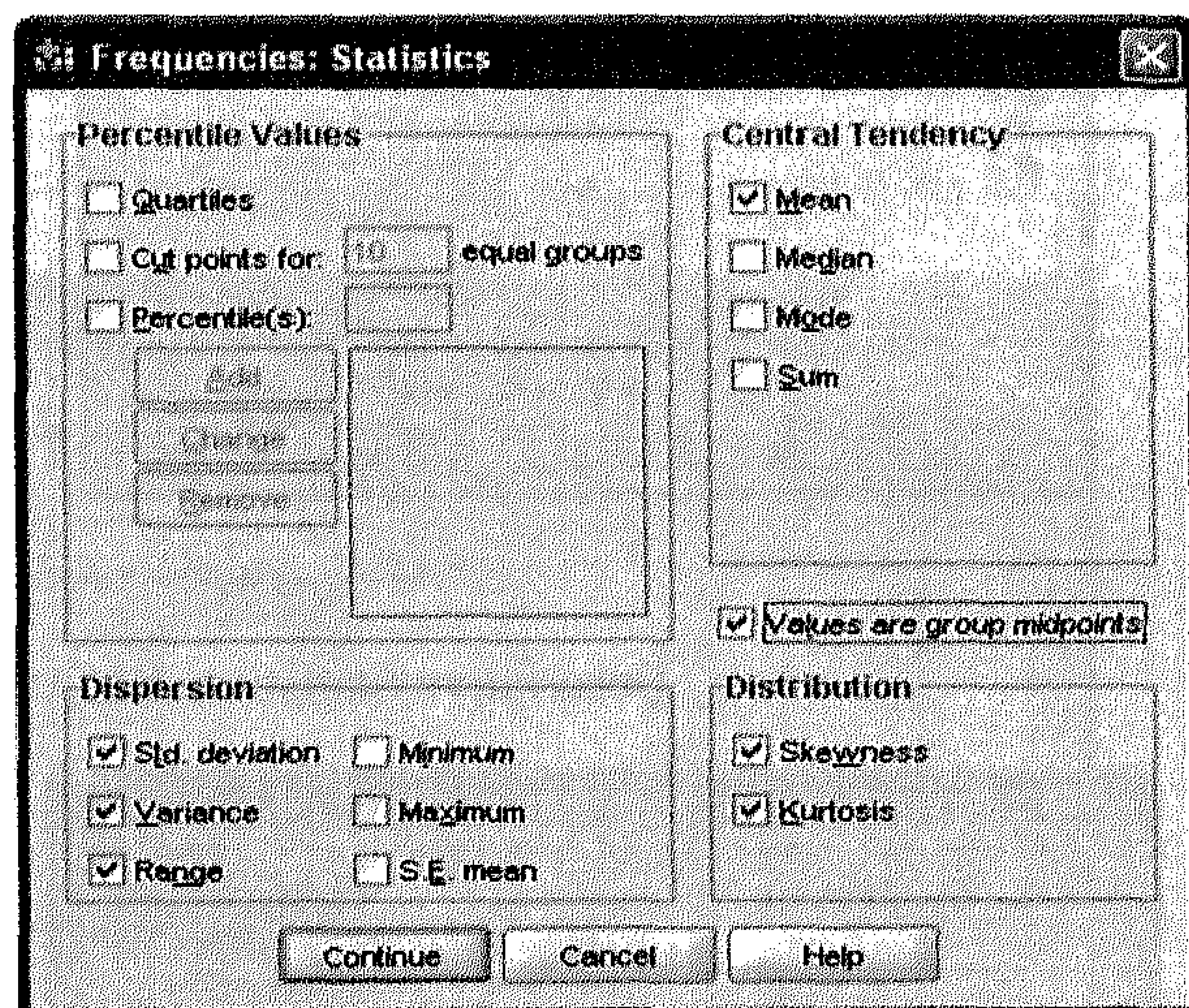
- من قائمة Analyze نختار Descriptive Statistics



- من القائمة المنسدلة نختار Frequencies
- تظهر شاشة جديدة بعنوان Frequencies ننقل المتغير x لقائمة Variables



- نضغط على الأمر Statistics فتظهر شاشة جديدة بعنوان Frequencies: Statistics نحدد الاختيار Values are group midpoints



- من قائمة Central Tendency نختار Mean،
- من قائمة Dispersion نختار Range, Std. deviation, Variance,
- ومن قائمة Distribution نختار Skewness, Kurtosis

- ثم نضغط على Continue فنعود للشاشة السابقة
- ثم نضغط على Ok فتظهر النتائج التالية:

Statistics

x

N	Valid	100
	Missing	0
Mean		4.2050
Std. Deviation		.15333
Variance		.024
Skewness		-.336
Std. Error of Skewness		.241
Kurtosis		-.495
Std. Error of Kurtosis		.478
Range		.60

من الجدول السابق نجد أن: عدد القيم $n = 100$، الوسط الحسابي للبيانات $\bar{x} = 4.205$، الانحراف المعياري $s = 0.15333$ والتباين $s^2 = 0.024$

ومن خلال قيمة الوسط والانحراف يمكن تعيين معامل الاختلاف:

$$c.v. = \frac{s}{\bar{x}} \times 100 = \frac{0.15333}{4.205} \times 100 = 3.646\%$$

ونجد أيضا أن معامل الالتواء $sk = -0.336$ وحيث أن قيمته سالبة فإن المنحنى ملتو ناحية اليسار، معامل التفرطح بالنسبة للتوزيع الطبيعي $k - 3 = -0.495$ ونجد أن منحنى التوزيع مفرطح. مدى البيانات هو $Range = 0.6$

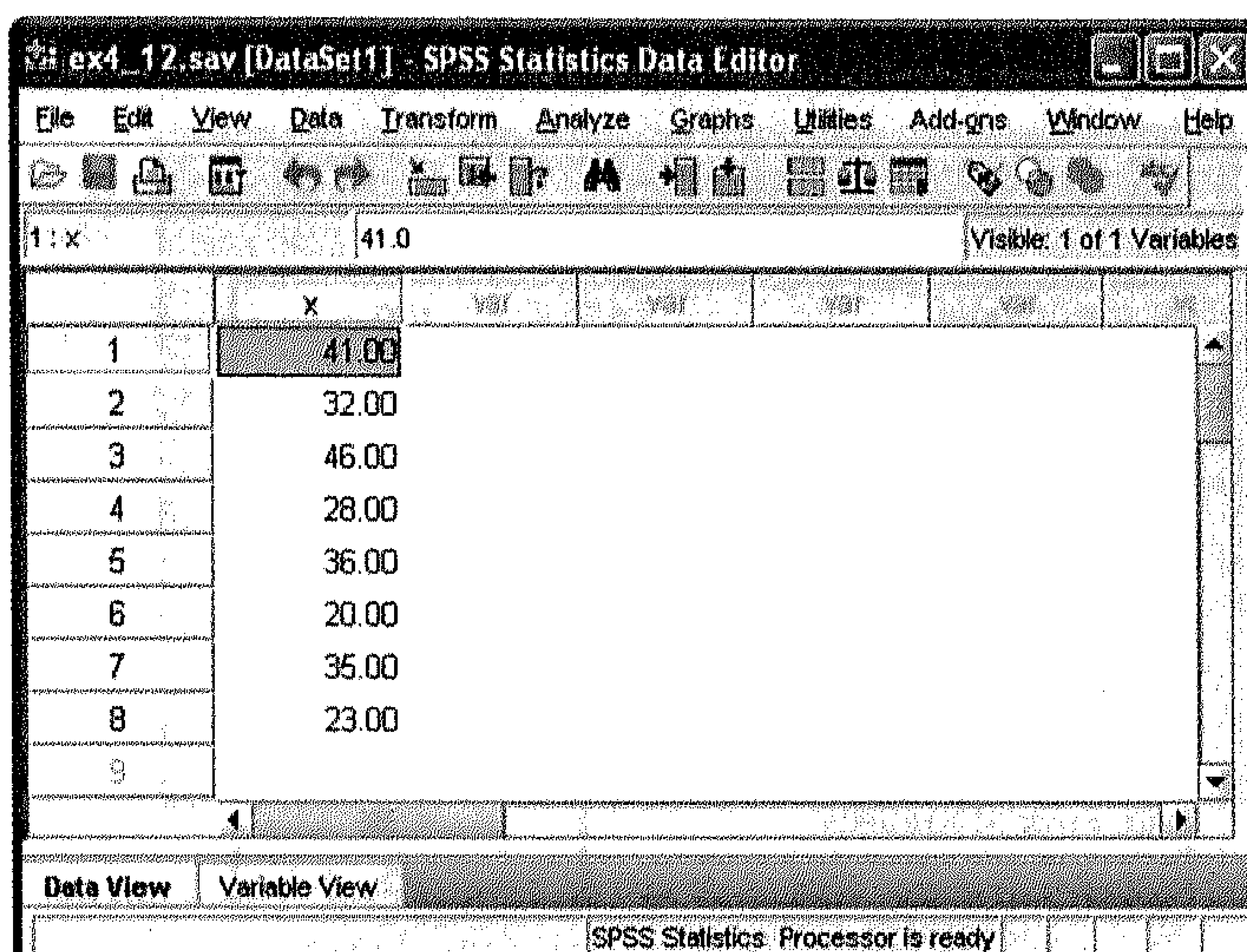
تطبيق (٤-٥)

من مثال (٤-١١): أوجد معامل الالتواء مع تحديد نوعه للبيانات التالية 41, 32, 46, 28, 36, 20, 35, 23 ثم عين معامل التفرطح.

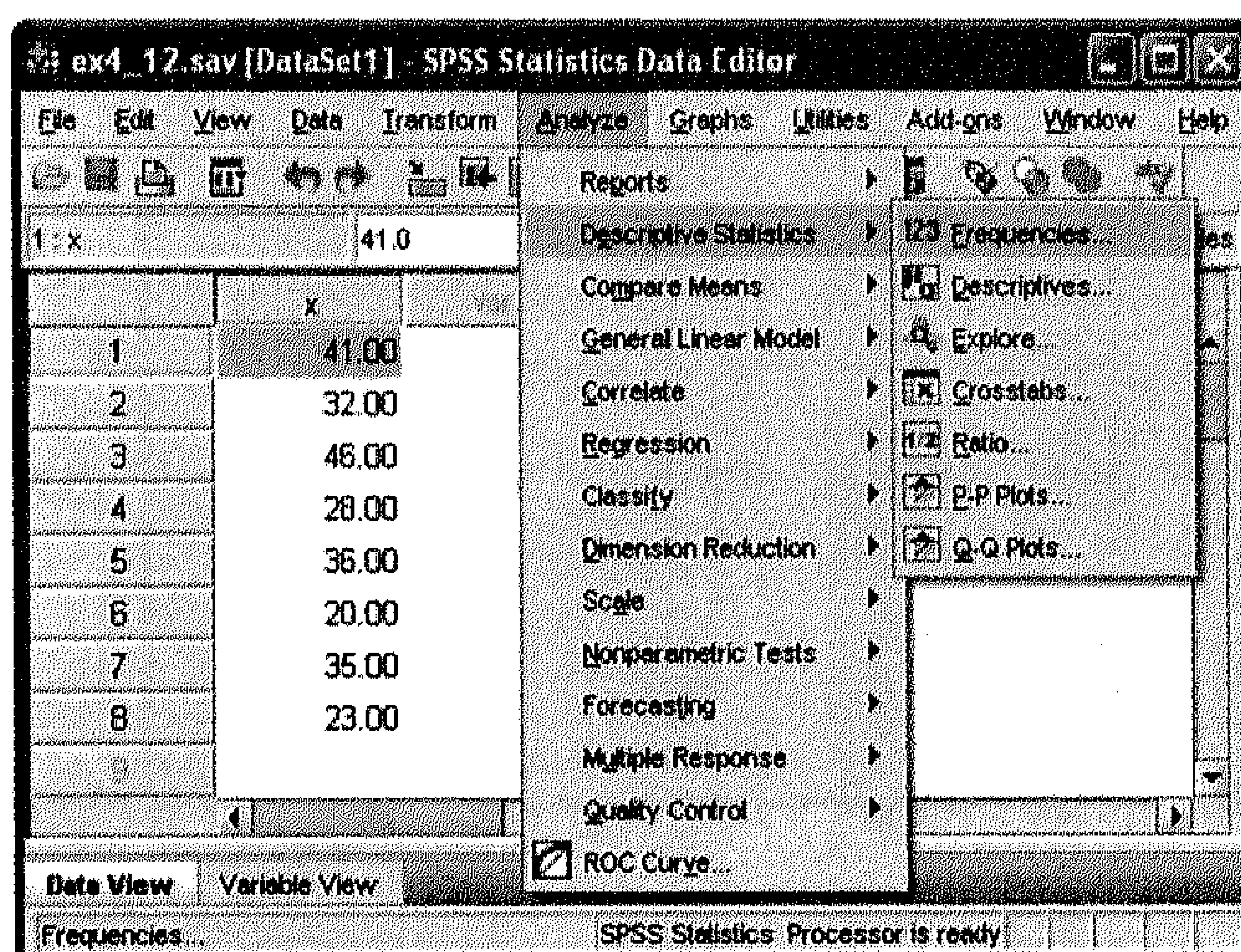
الحل:

١- نعرف متغير X في نافذة Variable View

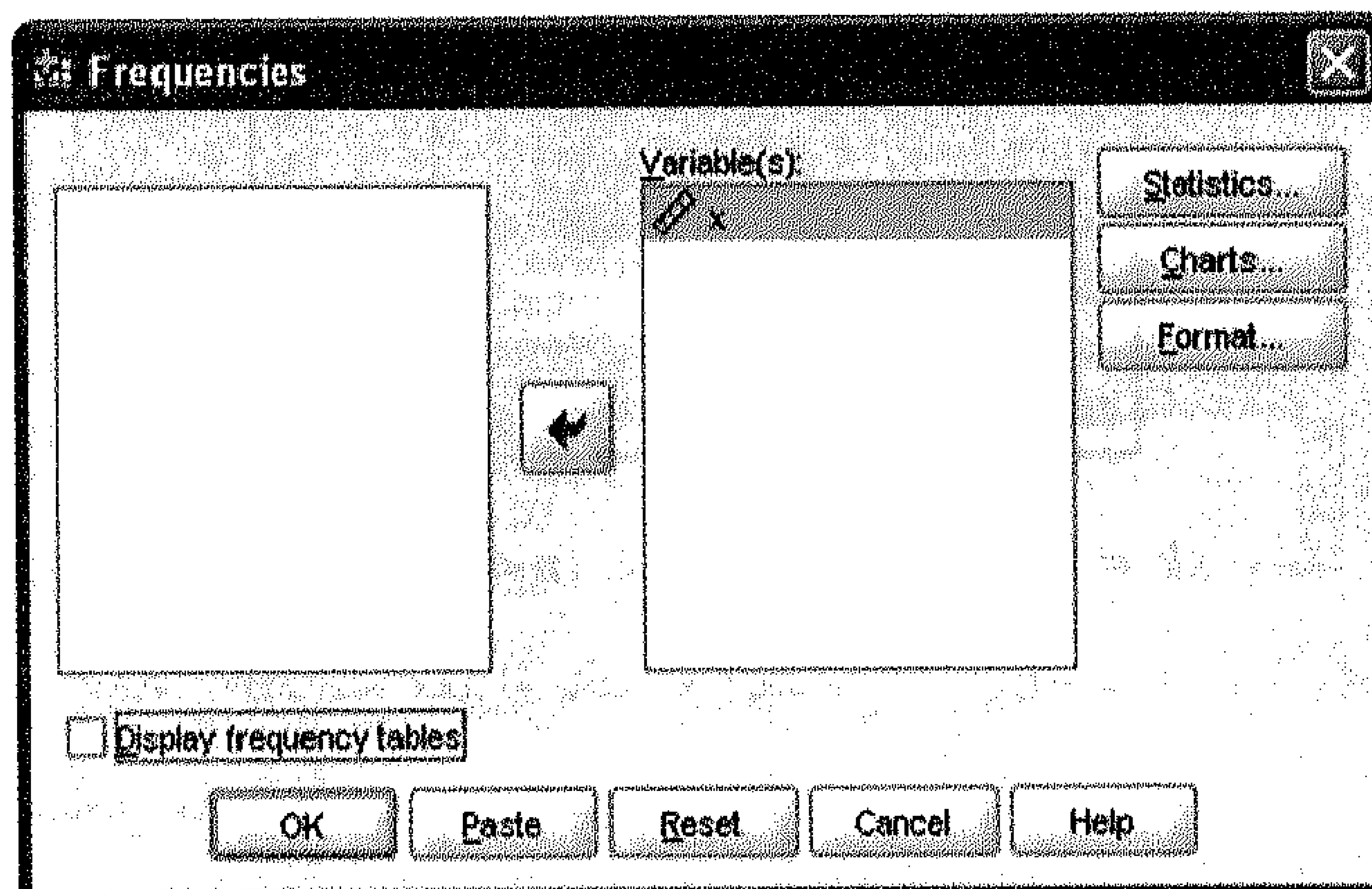
٢- ثم ننتقل لنافذة Data View وندخل البيانات للمتغير:



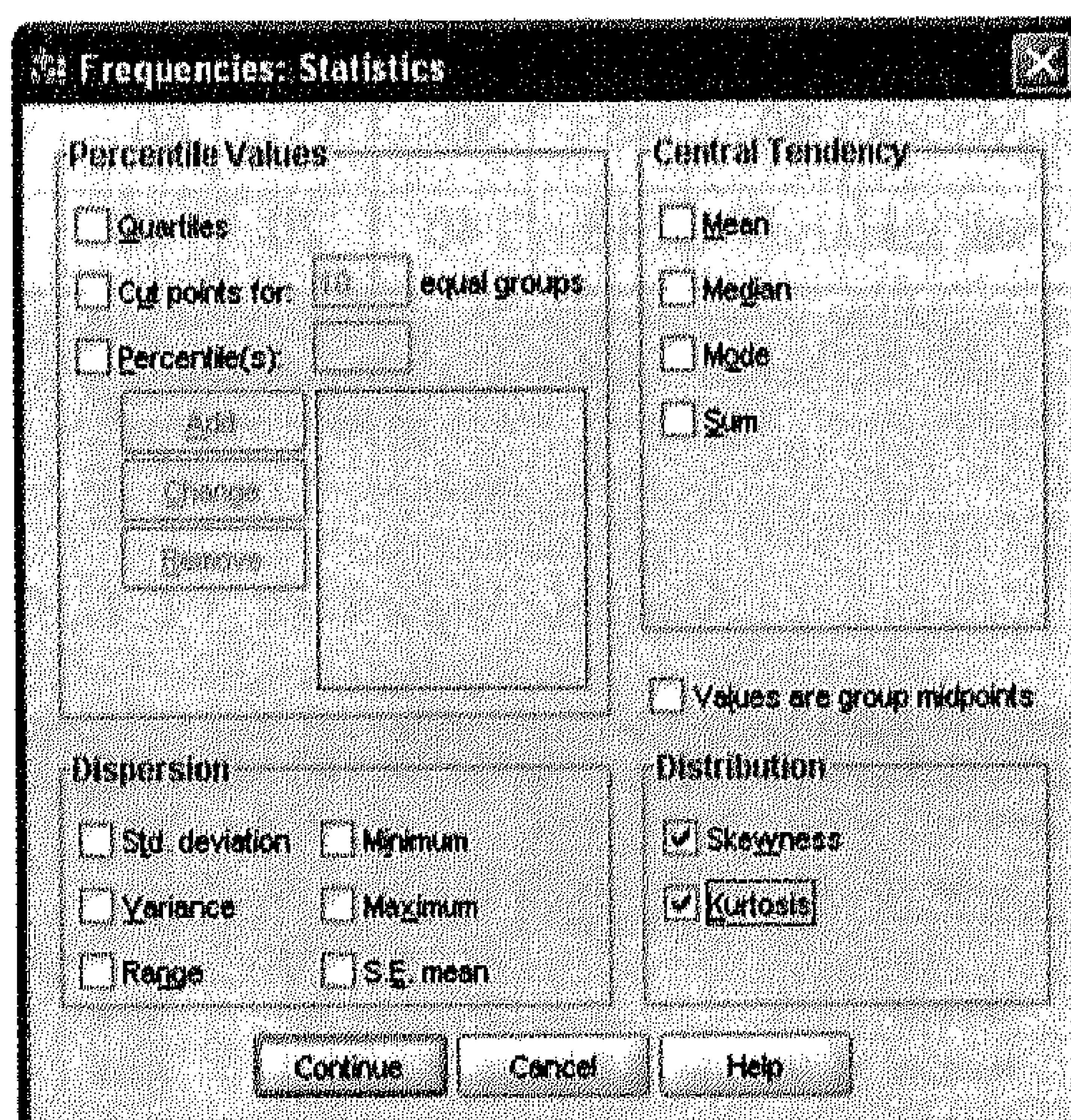
٣- من قائمة Analyze نختار الأمر Descriptive Statistics ومن القائمة المنسدلة نختار الأمر Frequencies



٤- فتظهر نافذة جديدة بعنوان Frequencies ننقل المتغير x لقائمة Variable(s)



٥- نضغط على الأمر Statistics فتظهر نافذة جديدة بعنوان Frequencies: Statistics



٦- نحدد كلاً من Skewness, Kurtosis من قائمة Distribution ثم نضغط على Continue فنعود للشاشة السابقة.

٧- نختار Ok فتظهر النتائج التالية:

Statistics

x

N	Valid	8
	Missing	0
Skewness		.008
Std. Error of Skewness		.752
Kurtosis		-.777
Std. Error of Kurtosis		1.481

ومن الجدول السابق فإن عدد القيم $n = 8$، معامل الالتواء $sk = 0.008$ ونجد أنه موجب لذا فإن الالتواء ناحية اليمين، معامل التفرطح مقارنة بالتوزيع الطبيعي $k - 3 = -0.777$ ونجد أنه سالب أي أنه أقل من 3 لذا فإن المنحنى مفرطح.

تطبيق (٤-٦)

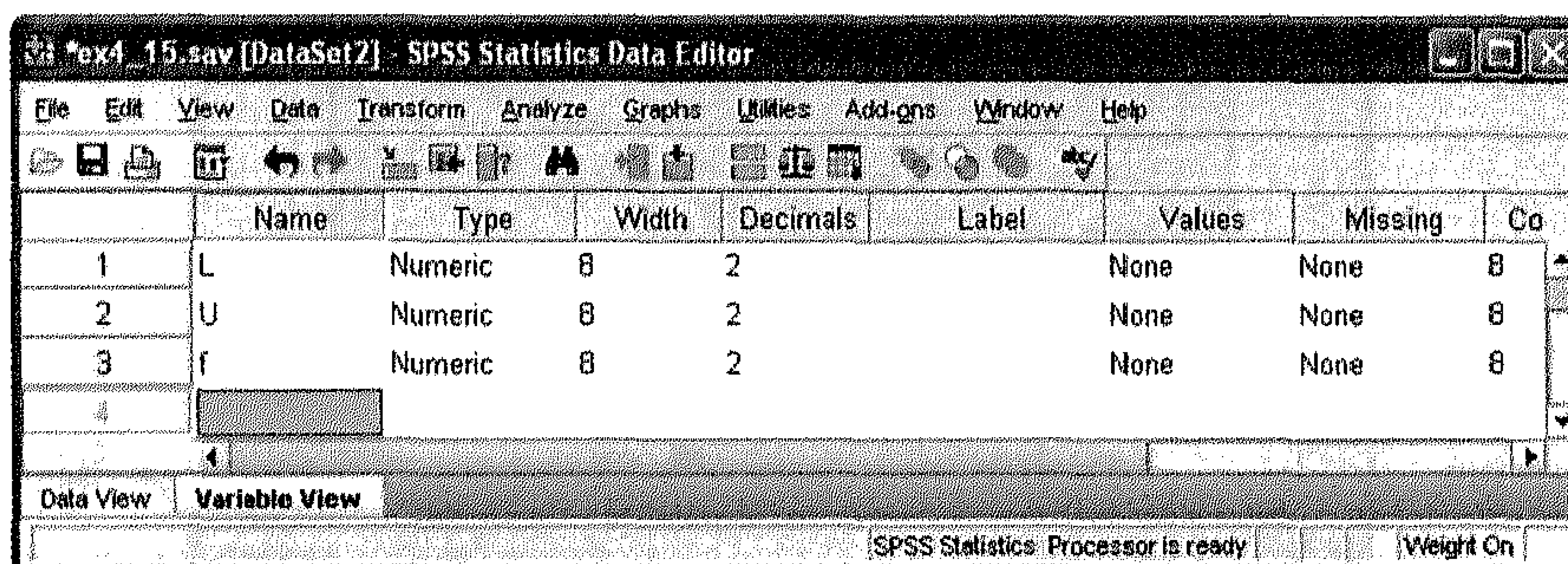
في مثال (٤-١٤):البيانات التالية تمثل أوزان مائة طالب مبوبة في الجدول التالي:

فترات الأوزان	40 –	45 –	50 –	55 –	60 –	65 –
عدد الطلاب	7	18	40	20	10	5

احسب قيمة معامل الالتواء والتفلطح وبين نوع منحنى التوزيع.

الحل:

١- نقوم أولاً بتعريف المتغيرات L, U, f في نافذة Variable View



	Name	Type	Width	Decimals	Label	Values	Missing	Co
1	L	Numeric	8	2		None	None	8
2	U	Numeric	8	2		None	None	8
3	f	Numeric	8	2		None	None	8

٢- ندخل بيانات الحدود الدنيا للفترات في المتغير L والحدود العليا في المتغير U والتكرار f

*ex4_15.sav [DataSet2] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : L 40.0 Visible: 3 of 3 Variables

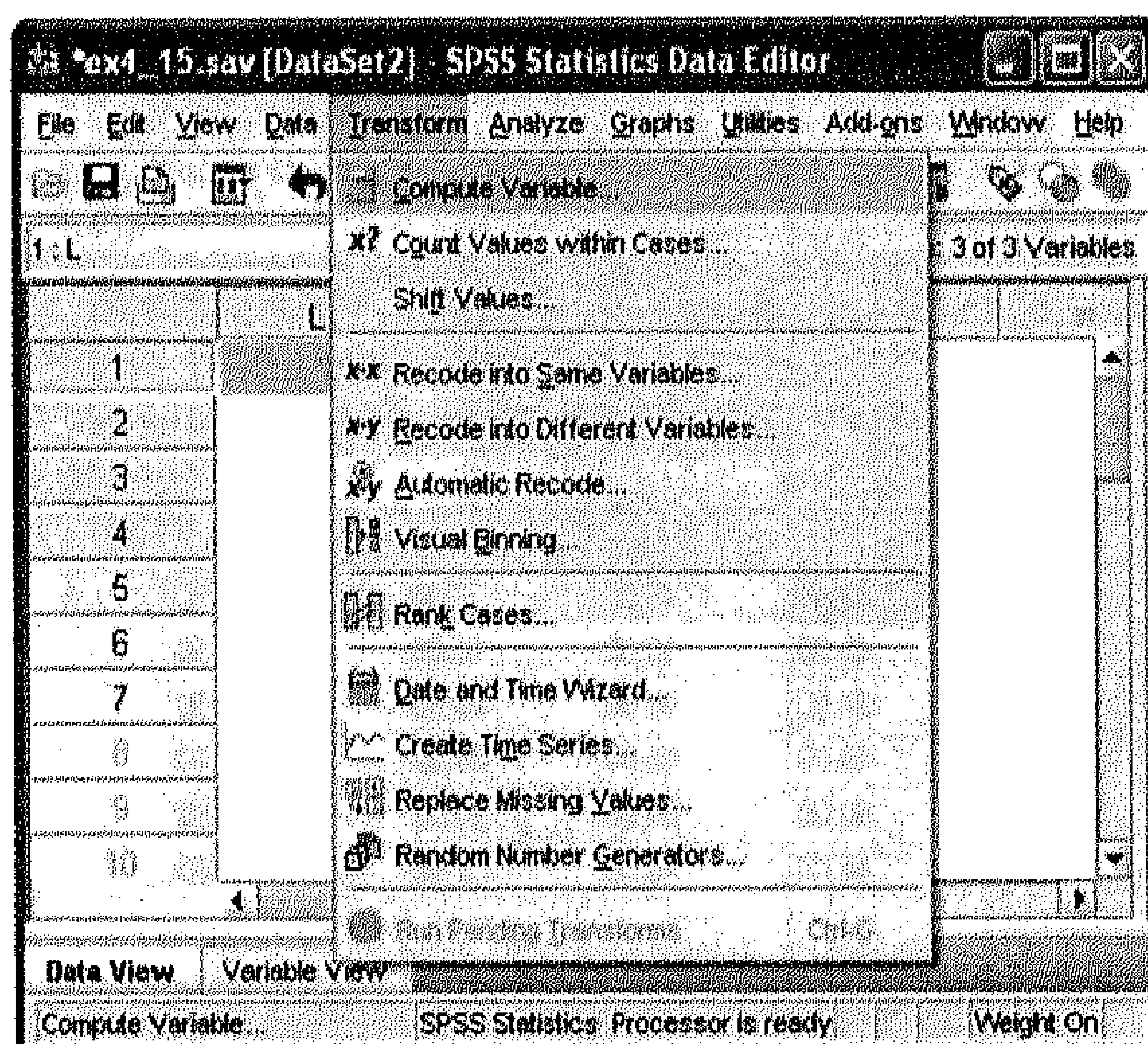
	L	U	f	var
1	40.00	45.00	7.00	
2	45.00	50.00	18.00	
3	50.00	55.00	40.00	
4	55.00	60.00	20.00	
5	60.00	65.00	10.00	
6	65.00	70.00	5.00	
7	.	.	.	

Data View | Variable View

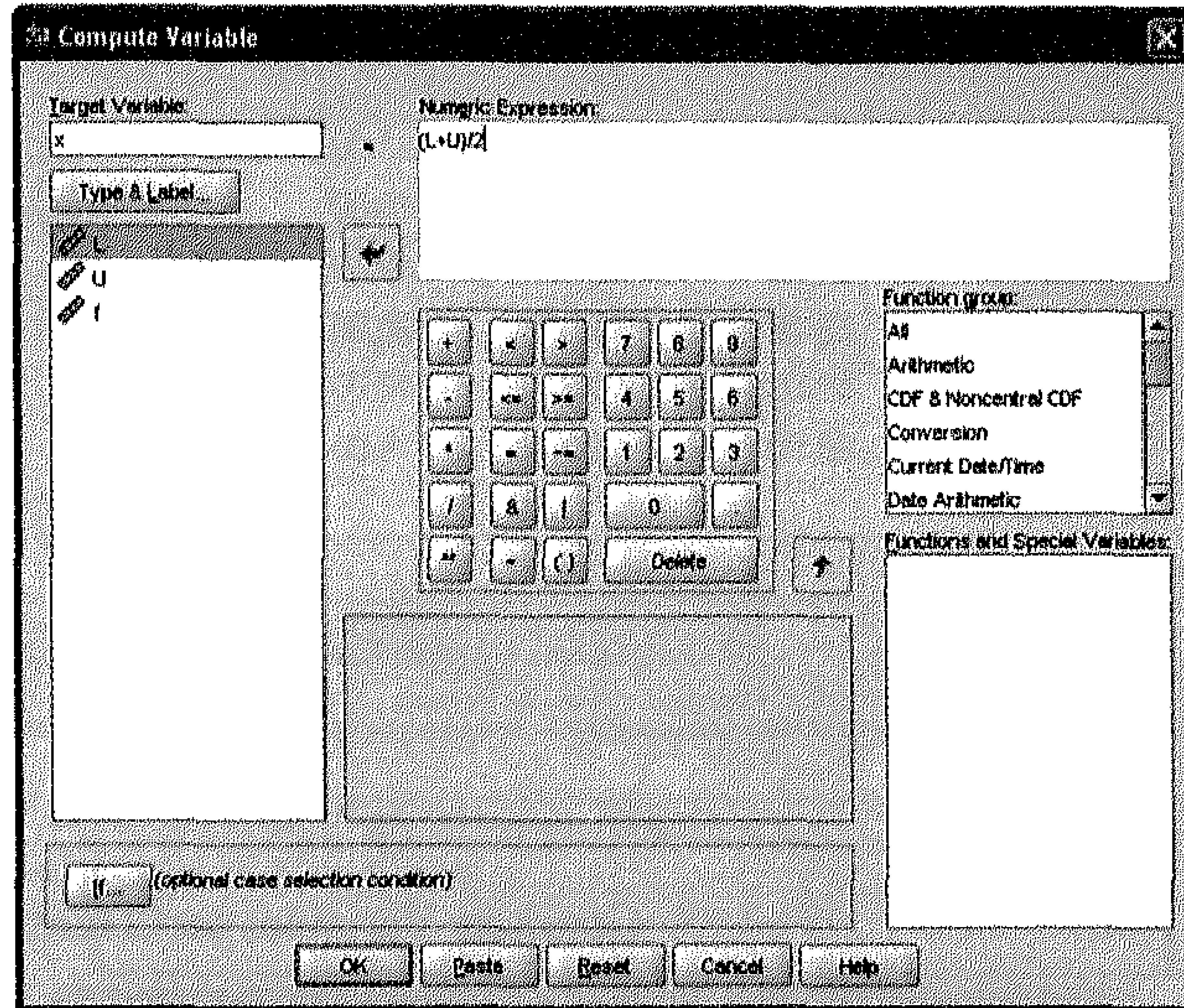
SPSS Statistics Processor is ready | Weight On

٣- نقوم بحساب مركز الفترات x وذلك باتباع الخطوات التالية:

- من قائمة Transform نختار Compute Variable



- تظهر نافذة جديدة بعنوان Compute Variable نكتب اسم المتغير الجديد وهو مركز الفترات x في خانة Target Variable ونكتب التعبير الرياضي الذي سيتم تعيين x باستخدامه في خانة Numeric Expression وهو $\frac{L+u}{2}$



- نختار Ok فنعود لملف البيانات وقد أضاف متغيراً جديداً x وهو مركز الفترات.

*ex4_15.sav [DataSet2] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : L 40.0 Visible: 4 of 4 Variables

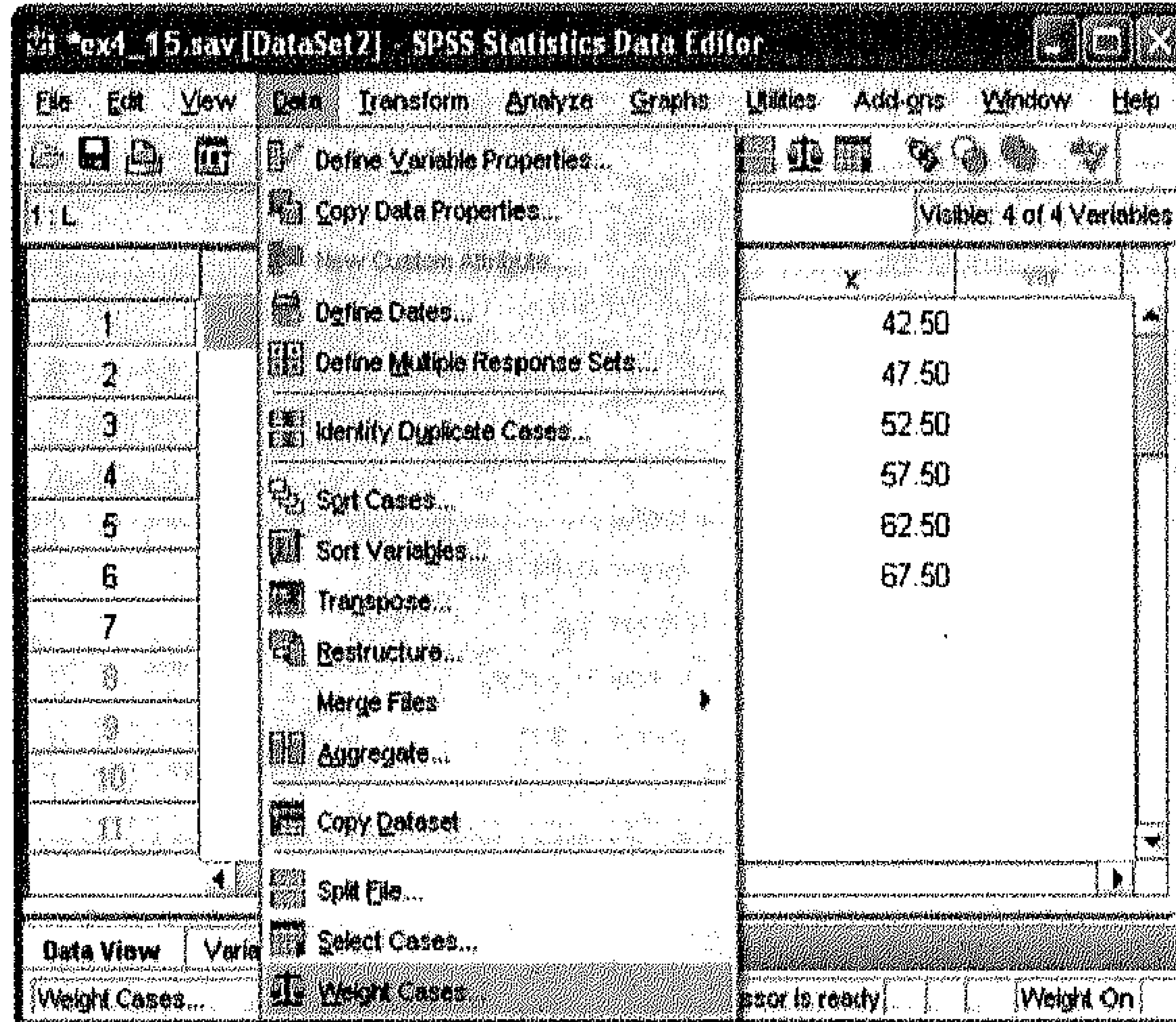
	L	U	f	x
1	40.00	45.00	7.00	42.50
2	45.00	50.00	16.00	47.50
3	50.00	55.00	40.00	52.50
4	55.00	60.00	20.00	57.50
5	60.00	65.00	10.00	62.50
6	65.00	70.00	5.00	67.50
7	.	.	.	.

Data View Variable View

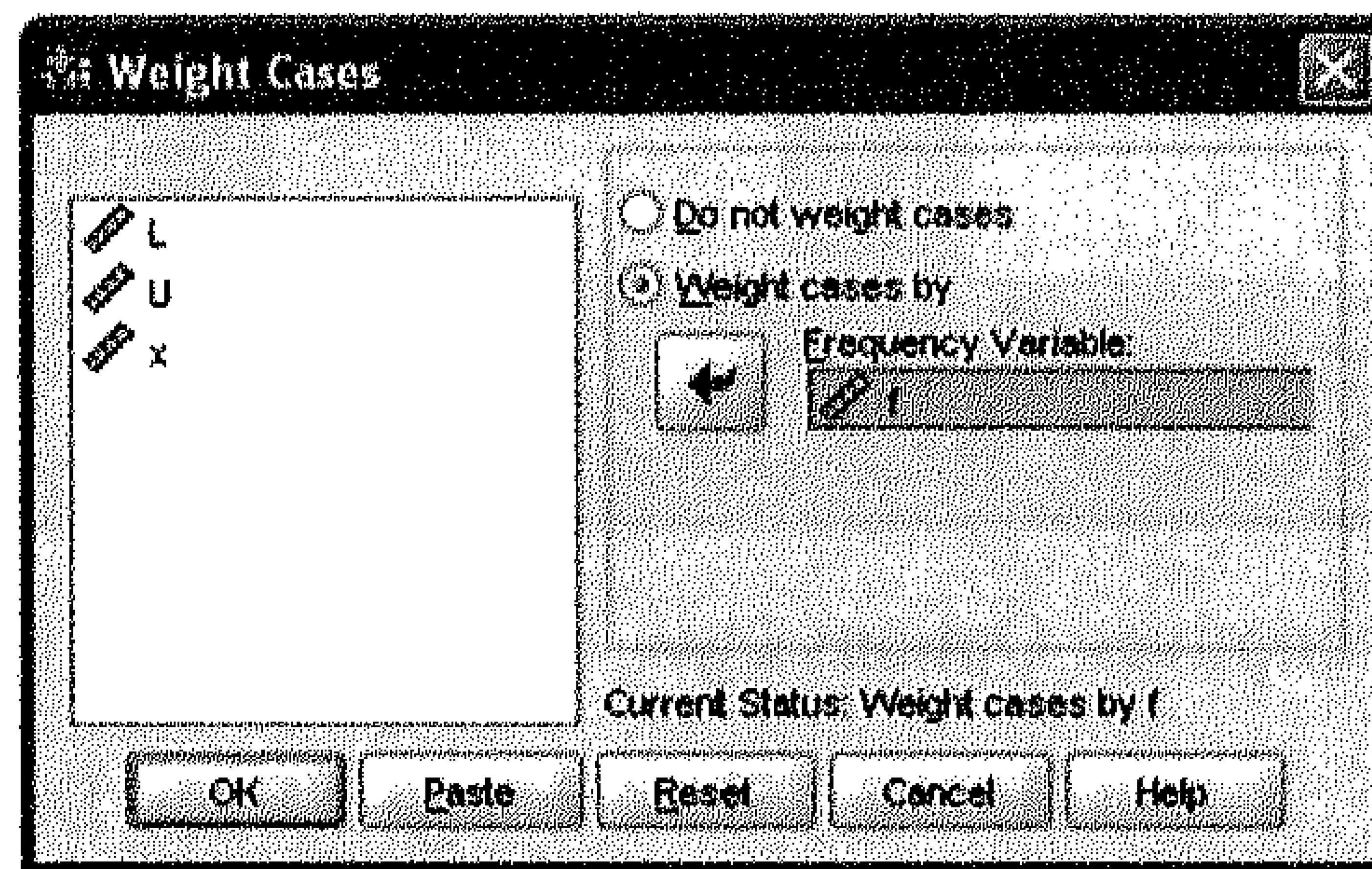
SPSS Statistics Processor is ready Weight On

٤- نقوم بتعريف f على أنه تكرار للفترات وذلك باتباع الخطوات التالية:

- من قائمة Data نختار Weight Cases



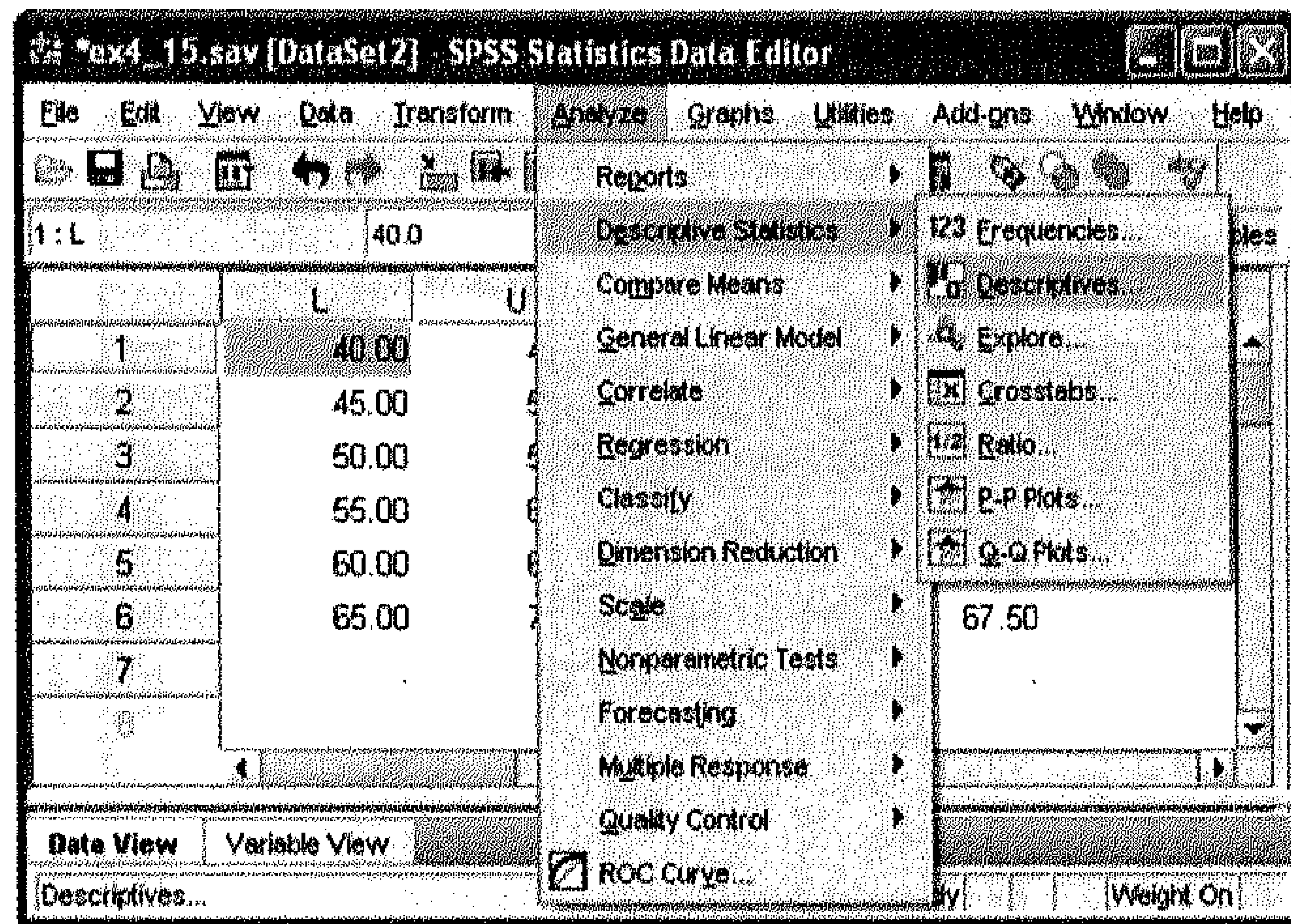
- تظهر نافذة جديدة بعنوان Weight Cases نختار Weight Cases by
- ننقل المتغير f لقائمة Frequency Variable



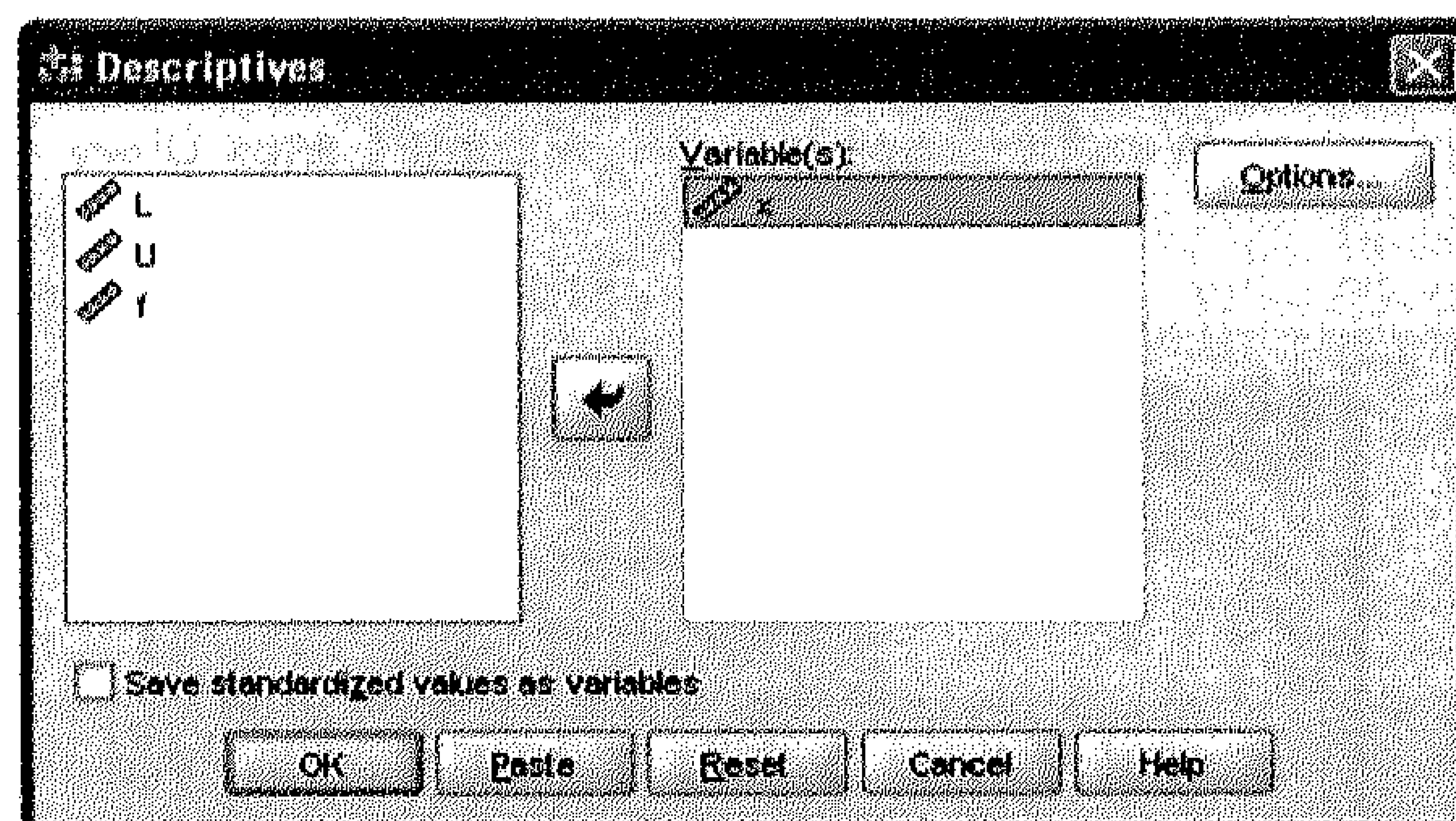
- نضغط على Ok فنعود لملف البيانات.

٥- لحساب قيمة معامل الالتواء والتفرطح نتبع الخطوات التالية:

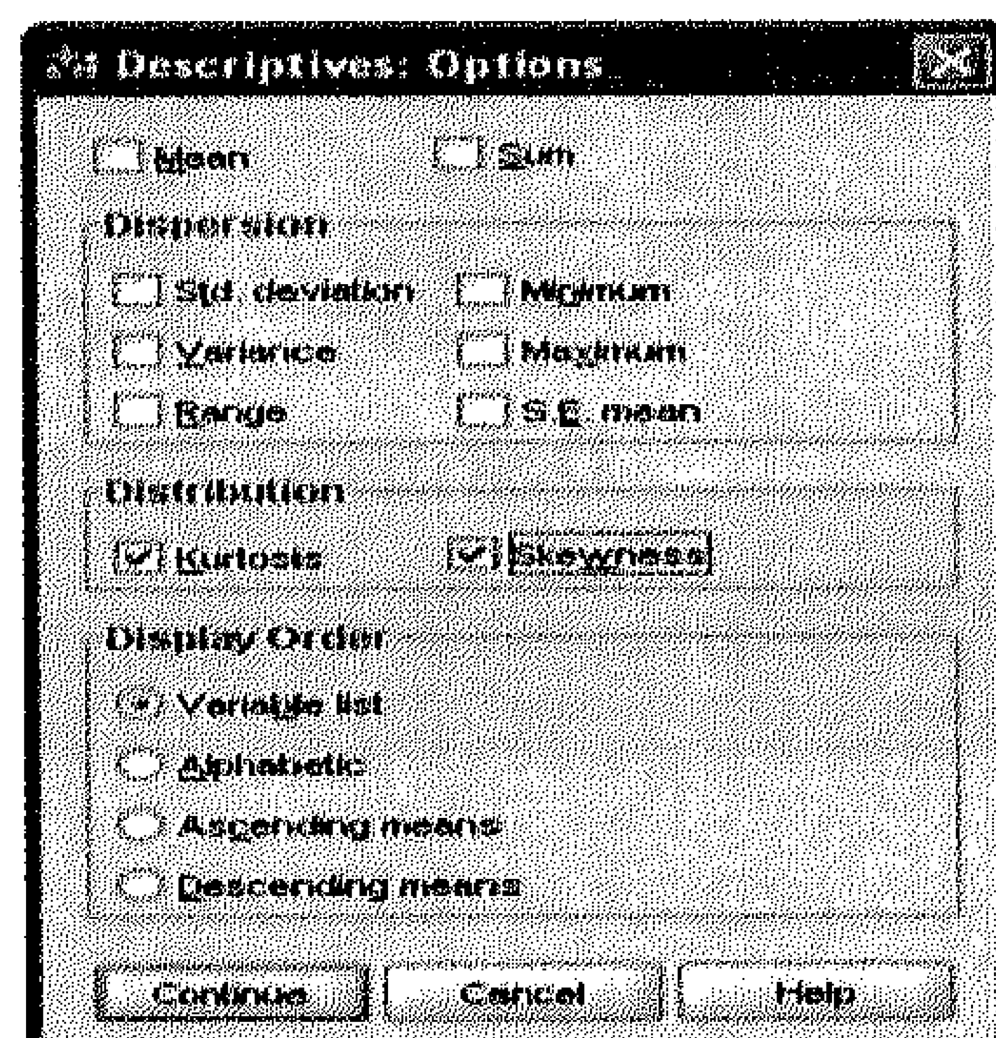
- من قائمة Analyze نختار Descriptive Statistics
- تظهر قائمة منسدلة نختار منها Descriptives



- تظهر نافذة جديدة بعنوان Descriptives ننقل المتغير x لقائمة Variable(s)



- نضغط على Options فتظهر شاشة جديدة بعنوان Descriptives: Options
- من قائمة Distribution نحدد كلاً من Kurtosis, Skewness



- بالضغط على Continue فنعود للشاشة السابقة.
- نضغط على Ok فتظهر النتائج التالية:

Descriptive Statistics

	N	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Std. Error
x	100	.342	.241	-.051	.478
Valid N (listwise)	100				

من الجدول السابق نجد أن معامل الالتواء هو $sk = 0.342$ وقيمته موجبه لذا فان المنحنى ملتويا ناحية اليمين، معامل التفرطح مقارنة بالتوزيع الطبيعي أي $k - 3 = -0.051$ وقيمته سالبة أي أقل من 3 لذا فإن المنحنى مفرطح.

أسئلة وتمارين (٤)

١- اشرح المقصود بمقاييس التشتت.

٢- متى نستخدم مقياس التشتت النسبي؟

٣- عرف معامل الاختلاف.

٤- ما الفرق بين معامل الاختلاف والقيم المعيارية؟

٥- أوجد الانحراف المعياري للأعداد 1, 2, 3,... , n

٦- ما المقصود بالالتواء؟ اذكر أنواع الالتواء.

٧- ما المقصود بالدرجة المعيارية؟

٨- كيف يتم مقارنة درجتي طالب في مادتين مختلفتين؟

٩- ما المقصود بالتفرطح؟

١٠- متي يقال إن المنحني ملتو ناحية اليمين؟

١١- ما الفرق بين الالتواء والتفلطح وفيم يستخدم كل منهما وكيفية حسابهما؟

١٢- ما الشروط الواجب توافرها لاستخدام التباين لمقارنة تجانس مجموعتين من البيانات؟

١٣- إذا كان الوسط الحسابي لمجموعة من البيانات 40 والانحراف المعياري 3 فما القيمة الأصلية إذا كانت قيمتها المعيارية 2.5؟

١٤- إذا كان الانحراف المعياري لمجموعة من القيم يساوى 4 وعدلت جميع هذه القيم حسب المعادلة $y = 3x + 8$ حيث y تمثل العلامة بعد التعديل، x العلامة قبل التعديل، فإن الانحراف المعياري بعد التعديل يصبح ....

١٥- إذا كان الوسط الحسابي لعلامات شعبة ما في مقرر الإحصاء يساوى 70 والتباين لها هو 36 فإن القيمة المعيارية للعلامة 60 هي ....

١٦- التباين للمشاهدات 5, 5, 5, 5, 5, 5, 5, 5, 5 هو ....

١٧- إذا كان التباين لمجموعة من البيانات هو 16 فإذا أضفنا القيمة 3 لجميع البيانات فإن التباين للبيانات بعد التعديل هو ....

١٨- معامل الاختلاف للبيانات 2, 2, 2, 2, 2, 2 يساوي..............................

١٩- إذا كان الوسط الحسابي لمجموعة من القيم 60 والتباين 16 فإن القيمة المعيارية للقيمة 92 هي ....

٢٠- إذا كان المئين الـ 75 يساوى 90 والمئين الـ 50 يساوى 77 والمئين الـ 25 يساوى 50 فإن معامل الالتواء يساوى .... ونصف المدى الربيعي يساوى ....

٢١- إذا كان الوسط الحسابي لمجموعة من البيانات يساوى 34 والوسيط لها يساوى 37 فإن منحنى البيانات ملتو ناحية............

٢٢- إذا كان الوسط الحسابي لمجموعتين من البيانات مختلف فإن .... يستخدم لمقارنة المجموعتين معا.

٢٣- معامل الالتواء للبيانات 3, 3, 3, 3, 3, 3, 3, 3, 3 يساوي ....

٢٤- إذا كان الوسط الحسابي لمجموعة من البيانات يساوي الوسيط فإن منحني البيانات ....

٢٥- فيما يلي التوزيع التكراري لأعمار 100مصباح كهربائي (بالساعة)

فئات العمر	0–	250–	500–	750–	1000–	1250–	1500–	1750–
عدد المصابيح	1	3	7	12	25	39	11	2

(أ) احسب معامل الاختلاف.

(ب) ادرس تماثل المنحنى.

٢٦- إذا كان لدينا البيانات التالية والتي تمثل درجات مجموعة من الطلاب في مقرر 1040 احص وكانت الدرجة النهائية من 100 وهى 63 ,71 ,96 ,74 ,63 ,60 ,72 ,88 ,87 ,75 ,93 ,84 ,77 ,71 ,91 ,89 فأوجد

(أ) المدى لدرجات الطلاب.

(ب) معامل الاختلاف.

(ج) هل منحنى البيانات مفرطح ؟

٢٧- فيما يلي التوزيع التكراري لأعمار مجموعة من البطاريات التي تستخدم في تشغيل الآلات الحاسبة الصغيرة (بالساعة):

فئات العمر	14–	614–	1214–	1814–	2414–	3014–
عدد البطاريات	23	7	9	2	4	5

فأوجد

(أ) معامل الالتواء باستخدام الربيعات.

(ب) الوسط الحسابي والانحراف المعياري.

(ج) معامل الاختلاف.

٢٨- فيما يلي التوزيع التكراري لدرجات الحرارة المئوية خلال 50يوماً متتالية

درجات الحرارة	16–	18–	20–	22–	24–	26–	28–
عدد الأيام	3	6	8	16	7	6	4

فاحسب

(أ) نصف المدى الربيع والانحراف المعياري.

(ب) معامل الاختلاف.

(ج) معامل الالتواء.

٢٩- الجدول التكراري الآتي يبين بيانات أعمار 30 مريضاً راجعوا أحد المستشفيات

الفترات	12 – 14	15 – 17	18 – 20	21 – 23	24 – 26	27 – 29
التكرار	8	4	7	6	2	3

والمطلوب

(أ) حساب المدى لأعمار المرضى.

(ب) ادرس تماثل منحنى البيانات.

(ج) حساب معامل التفرطح.

٣٠- البيانات التالية تمثل أوزان 100 طالباً مبوبة كما يلي:

فترات الوزن	40 – 44	45 – 49	50 – 54	55 – 59	60 – 64	65 – 69
عدد الطلاب	7	18	40	20	10	5

أوجد

(أ) نصف المدى الربيعي لأوزان الطلاب.

(ب) التباين للأوزان.

(ج) معامل الاختلاف.

٣١- إذا كان متوسط دخل مجموعة من الأطباء بالمستشفى الجامعي هو 3500 ريالاً سعودياً بانحراف معياري قدره 600 ريال فأوجد نسبة الأطباء الذين يتراوح دخلهم بين 4370، 2630 ريالاً.

٣٢- إذا كان لدينا ثلاث عينات أحجامها على التوالي هي $n_1 = 15, n_2 = 20, n_3 = 25$ وكانت تبايناتها هي $s_1^2 = 9, s_2^2 = 8.5, s_3^2 = 6.4$ فأوجد الانحراف المعياري للعينة الناتجة من دمج تلك العينات معاً.

٣٣- إذا كانت درجات مجموعتين من طلاب البرنامج الموحد للكليات الصحية في امتحان 106 إحص من 15 درجة كالتالي:

الدرجة	4	5	6	7	8	9	10
عدد طلاب المجموعة الأولى	2	8	13	35	21	16	2
عدد طلاب المجموعة الثانية	3	5	20	23	24	11	7

(أ) قارن بين درجات الطلاب في المجموعتين؟

(ب) إذا فرض إن البرنامج الموحد مكون من مجموعتين فقط فما هو الوسط الحسابي والتباين لدرجات طلاب البرنامج الموحد في مقرر 106 إحص؟

٣٤- إذا كان الوسط الحسابي والتباين لعلامات 30 طالبا في مقرر الإحصاء يساوى 9 ,60 وعدلت العلامات وفق المعادلة التالية $y = 80 - 0.25\,x$ حيث x تمثل العلامات قبل التعديل، y العلامات بعد التعديل فأوجد قيمة الوسط والتباين للدرجات بعد التعديل؟

٣٥- إذا كان ما لا يقل عن 20% من طلاب إحدى الشعب في مقرر 1040 إحص تتراوح درجاتهم بين 63, 79 بتباين مقداره 9 درجات فما هو متوسط درجات طلاب تلك الشعبة؟

٣٦- البيانات التالية تمثل رواتب 100 موظف في إحدى الشركات مبوبة كما في الجدول التالي:

الرواتب	70 – 79	80 – 89	90 – 99	100 – 109	110 – 119	120 – 129	130 – 139
عدد الموظفين	5	7	21	33	18	13	3

فأوجد

(أ) مدى الرواتب.

(ب) معامل التفرطح وبين نوعه.

(ج) معامل الاختلاف.

٣٧- إذا كان إنتاج 15 قطعة أرض متساوية المساحة من محصول القمح بالإردب كالتالي:

26	14	21	18	17	16	24	10	20	18	9	10	12	23	19

(أ) أوجد مدى الإنتاج لقطع الأرض من القمح

(ب) ادرس تماثل منحنى توزيع محصول القمح.

(ج) هل منحنى توزيع محصول القمح مفرطح؟

٣٨- إذا كان لدينا مجموعتين من البيانات معامل الاختلاف للمجموعة الأولى هو 2.33% ومعامل الاختلاف للمجموعة الثانية هو 0.642% فإذا قمنا بتعديل بيانات المجموعة الأولى تبعاً للعلاقة التالية $y_1 = 0.25\, x_1$ وتم تعديل بيانات المجموعة الثانية تبعاً للعلاقة التالية $y_2 = 4x_2$ فقارن بين المجموعتين بعد إجراء التعديل المقترح.

الفصل الخامس

الوصف الإحصائي لمتغيرين

TWO VARIABLES DESCRIPTION

٥-١ مقدمة

لقد تعرضنا في الفصول السابقة لبعض الطرق التي تستخدم لوصف بيانات متغير واحد X لظاهرة ما، وهي تكوين الجداول التكرارية، التمثيل البياني، تعيين بعض المقاييس الإحصائية. وسوف نخصص هذا الفصل للتعامل مع أكثر من متغير وسنركز على شرح كيفية التعامل مع متغيرين ووصفهما إحصائياً باستخدام:

١- التمثيل البياني.

٢- الجداول المزدوجة للبيانات.

٣- حساب بعض المقاييس الإحصائية.

وسوف نهتم بدراسة متغيرين فقط X, Y وسنقوم بدراسة العلاقة بين المتغيرين ووصفها إحصائيا وذلك بتكرار إجراء تجربة مناسبة عددا كافيا من المرات وليكن n ونحصل على عدد n من المشاهدات عادة ما تلخص في جدول على الصورة

X	x_1	x_2	...	x_n
Y	y_1	y_2	...	y_n

والتي يمكن أن تكتب كمجموعة نقاط أو أزواج

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

ففي كثير من النواحي المعملية نجد أنه قد توجد علاقة ما بين متغيرين (أو أكثر) فمثلا:

- أوزان الذكور البالغين تعتمد بدرجة معينة على أطوالهم.
- ضغط وزن معين من الغاز يعتمد على درجة حرارته وحجمه.
- استهلاك سيارة معينة للوقود يعتمد على المسافة التي تقطعها.

وفي بعض الأحيان يكون من المرغوب فيه التعبير عن هذه العلاقة بصورة رياضية وذلك بتحديد المعادلة التي تربط بين تلك المتغيرات. وبعد الحصول على مثل هذه العلاقة يصبح من السهل التنبؤ (إيجاد قيمة تقديرية) بما سيكون عليه أحد المتغيرات متى علمت قيمة المتغير الآخر. مثلا عن طريق إجراء عدة مشاهدات للمسافة المقطوعة بسيارة ما وكمية الوقود المستهلكة يمكننا إيجاد علاقة رياضية تربط بين هذين المتغيرين بحيث يصبح في الإمكان:

(أ) التنبؤ بكمية الوقود اللازمة لقطع مسافة معينة.

(ب) التنبؤ بالمسافة التي يمكن أن تقطعها السيارة بكمية معينة من الوقود.

٥-٢ شكل الانتشار (Scatter Diagram)

يمكن تمثيل البيانات المزدوجة للمتغيرين (X, Y) عن طريق رسم شكل الانتشار للبيانات وذلك برسم محورين X, Y ورسم نقاط طبقاً لنظام الإحداثيات المتعامدة يظهر شكل يسمى بشكل الانتشار (scatter diagram) للمشاهدات المعطاة. ويستخدم هذا الشكل لتحديد نوع العلاقة بين المتغيرين X,Y هل هي علاقة عكسية أم طرديه أو لا يوجد بينهما علاقة أم إنها علاقة خطية أو غير خطية.

ولكن رسم منحنى يمر بمعظم النقاط سيكون مختلفاً كل حسب نظرته الخاصة، فإن استخدام هذا المنحنى في التنبؤ سيؤدى إلى عدة نتائج متباينة ينقصها قدر كبير من الدقة. كما أنه تظهر صعوبة في تطبيق الطريقة السابقة إذا كنا ندرس العلاقة بين ثلاثة متغيرات (تكون نقاط الانتشار هي نقاط في الفراغ ذو الثلاث أبعاد) كذلك استحالتها في حالة دراسة العلاقة بين أكثر من ثلاثة متغيرات.

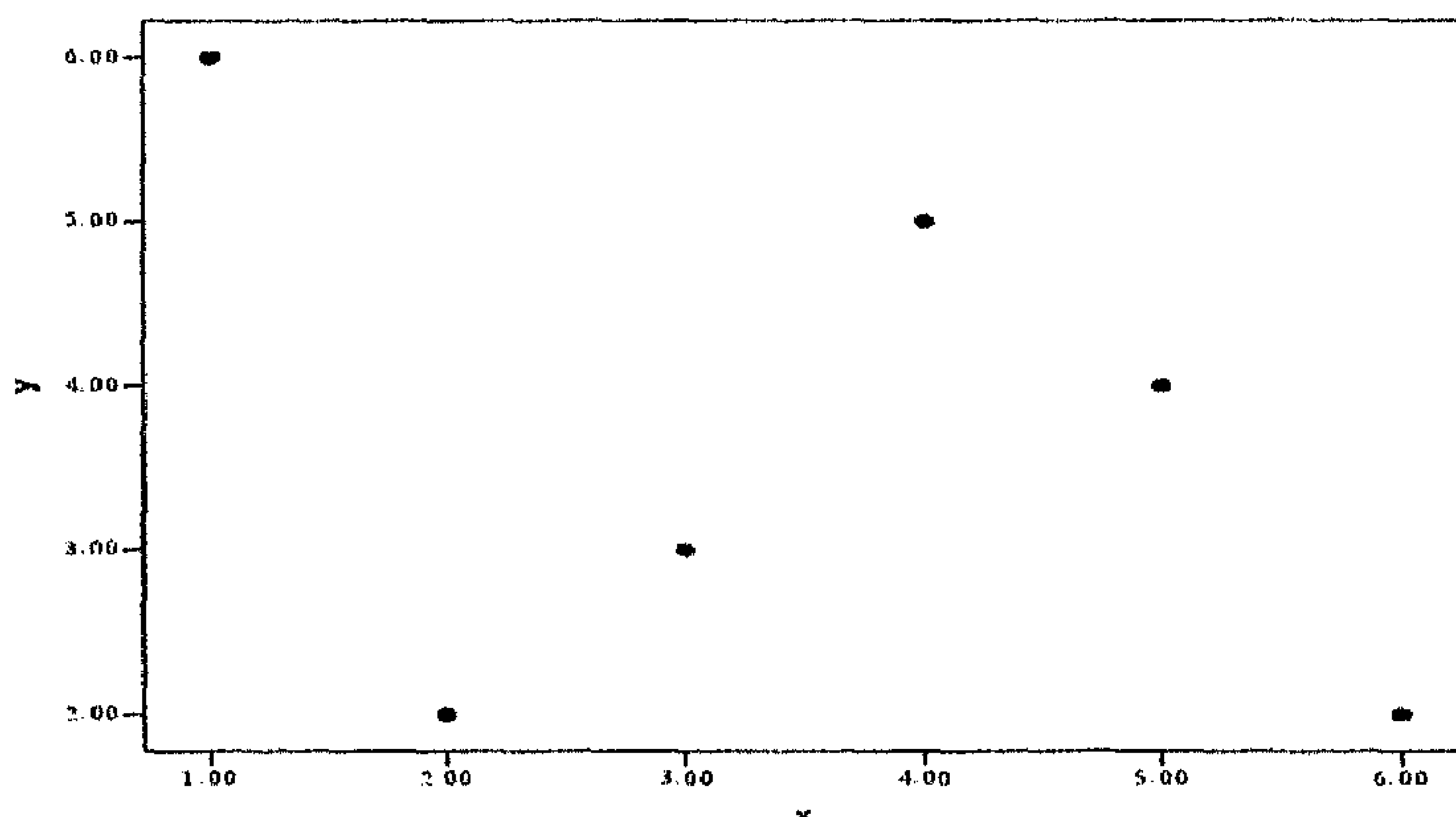
مثال (٥-١)

ارسم شكل الانتشار لمجموعة البيانات التالية:

X	2	4	1	3	5	6
Y	2	5	6	3	4	2

الحل:

برسم المحاور X, Y ورسم نقاط $(x_i, y_i),\ i = 1, 2, \dots 6$ سوف نحصل على شكل الانتشار التالي



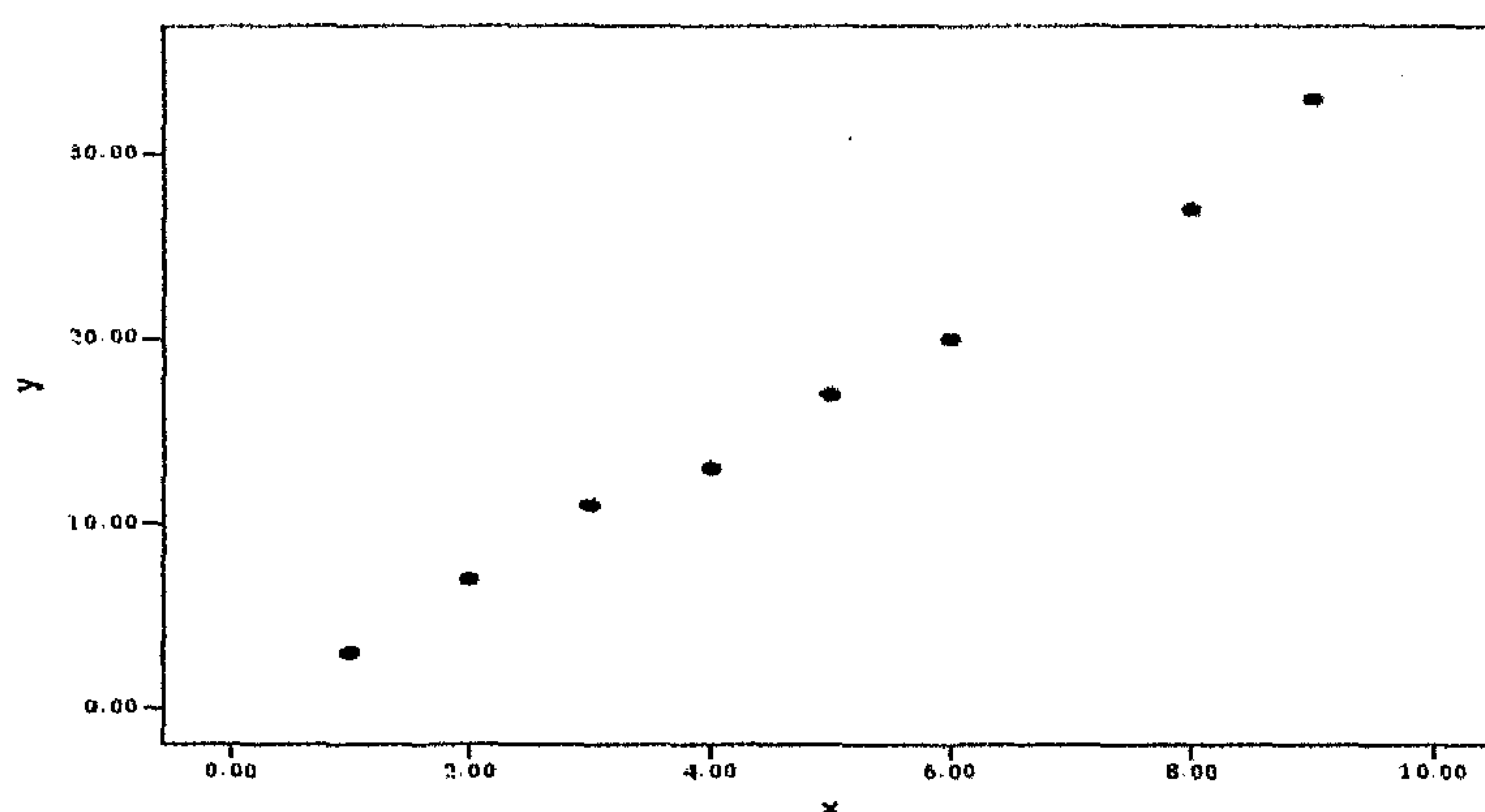
مثال (٥–٢)

ارسم شكل الانتشار لمجموعة البيانات التالية وحدد نوع العلاقة بينهما.

X	2	4	1	3	5	6	8	10
Y	7	13	3	11	17	20	27	33

الحل

برسم المحاور X, Y ورسم نقاط $(x_i, y_i),\ i = 1, 2, \dots 8$ سوف نحصل على شكل الانتشار التالي:



ونلاحظ من شكل الانتشار أن العلاقة بين المتغيرين X, Y هي علاقة خطية طرديه.

٥-٣ الجداول التكرارية المزدوجة (Frequency Cross Tables)

لقد تم في الفصل الثاني عرض البيانات لمتغير X باستخدام التوزيعات التكرارية وسوف نقوم بإنشاء جدول تكراري مماثل في حالة متغيرين X,Y كما بالمثال التالي.

مثال (٥-٣)

كون الجدول التكراري المزدوج لمتغيري النوع (X) ومستوى التعليم(Y)

النوع X	أنثى	أنثى	ذكر	ذكر	ذكر	أنثى	ذكر	أنثى	ذكر
التعليم Y	ضعيف	متوسط	عالي	عالي	متوسط	ضعيف	ضعيف	متوسط	عالي
النوع X	أنثى	أنثى	أنثى	ذكر	ذكر	ذكر	أنثى	ذكر	أنثى
التعليم Y	متوسط	ضعيف	متوسط	عالي	عالي	متوسط	ضعيف	ضعيف	متوسط
النوع X	ذكر	أنثى	أنثى	ذكر	أنثى	أنثى	ذكر		
التعليم Y	عالي	متوسط	ضعيف	متوسط	متوسط	عالي	ضعيف		

الحل:

المتغيرات هنا هي متغيرات وصفية حيث المتغير الأول هو النوع وله قيمتان فقط هما ذكر أو أنثى أما المتغير الثاني فهو التعليم وله ثلاث قيم هي ضعيف، متوسط، عالٍ. وسوف يكون الجدول التكراري المزدوج بالصورة التالية:

التعليم / النوع	ضعيف	متوسط	عالي	المجموع
أنثى	5	7	1	13
ذكر	3	3	6	12
المجموع	8	10	7	25

حيث القيم الموجودة بالجدول تمثل عدد مرات تكرار قيمة متغير التعليم ومتغير النوع معاً في نفس الوقت فعلى سبيل المثال القيمة (أنثى، ضعيف) تكررت 5 مرات في البيانات المعطاة، وأيضا (ذكر، ضعيف) تكررت 3 مرات في البيانات ونجد أن القيمة ضعيف تكررت 8 مرات بغض النظر عن النوع، وهى ناتج جمع العمود الأول ومن الجدول السابق يمكن تعيين التوزيع التكراري لكل متغير على حده والذي يسمي بالتوزيع التكراري الهامشي كالتالي:

- التوزيع التكراري الهامشي للنوع والذي يمثل العمود الأخير(مجموع كل صف)

النوع	ذكر	أنثى	المجموع
التكرار	12	13	25

- التوزيع التكراري الهامشي لمستوى التعليم والذي يمثل الصف الأخير (مجموع كل عمود)

مستوى التعليم	ضعيف	متوسط	عالي	المجموع
التكرار	8	10	7	25

مثال (٥-٤)

أراد أحد الباحثين دراسة العلاقة بين مستوى التعليم وعدد الأولاد فقام بأخذ عينة من 25 شخصاً من أحد الأحياء وقام بتوجيه بعض الأسئلة التي تتضمن مستوى التعليم للشخص وعدد الأولاد وكانت النتائج كالتالي:

مستوى التعليم X	ضعيف	متوسط	عالي	عالي	متوسط	ضعيف	ضعيف	متوسط	عالي
عدد الأولاد Y	4	3	5	4	2	4	3	2	5
مستوى التعليم X	متوسط	ضعيف	متوسط	عالي	عالي	متوسط	ضعيف	ضعيف	متوسط
عدد الأولاد Y	2	4	5	1	2	3	1	2	1
مستوى التعليم X	عالي	متوسط	ضعيف	متوسط	متوسط	عالي	ضعيف		
عدد الأولاد Y	1	2	3	4	5	1	3		

كون الجدول التكراري المزدوج لمستوى التعليم (X) وعدد الأولاد(Y) .

الحل:

المتغير الأول هو مستوى التعليم ويأخذ القيم ضعيف، متوسط، عالي وهو متغير نوعي.

المتغير الثاني هو عدد الأولاد في الأسرة ويأخذ القيم 5 ,4 ,3 ,2 ,1 وهو متغير كمي.

وبالتالي فان الجدول التكراري المزدوج للمتغيرين X, Y يمكن تكوينه كالتالي:

مستوى التعليم / عدد الأولاد	1	2	3	4	5	المجموع
ضعيف	1	1	3	3	0	8
متوسط	1	4	2	1	2	10
عالي	3	1	0	1	2	7
المجموع	5	6	5	5	4	25

التوزيع التكراري الهامشي لمستوى التعليم هو

مستوى التعليم	ضعيف	متوسط	عالي	المجموع
التكرار	8	10	7	25

التوزيع التكراري الهامشي لعدد الأولاد هو

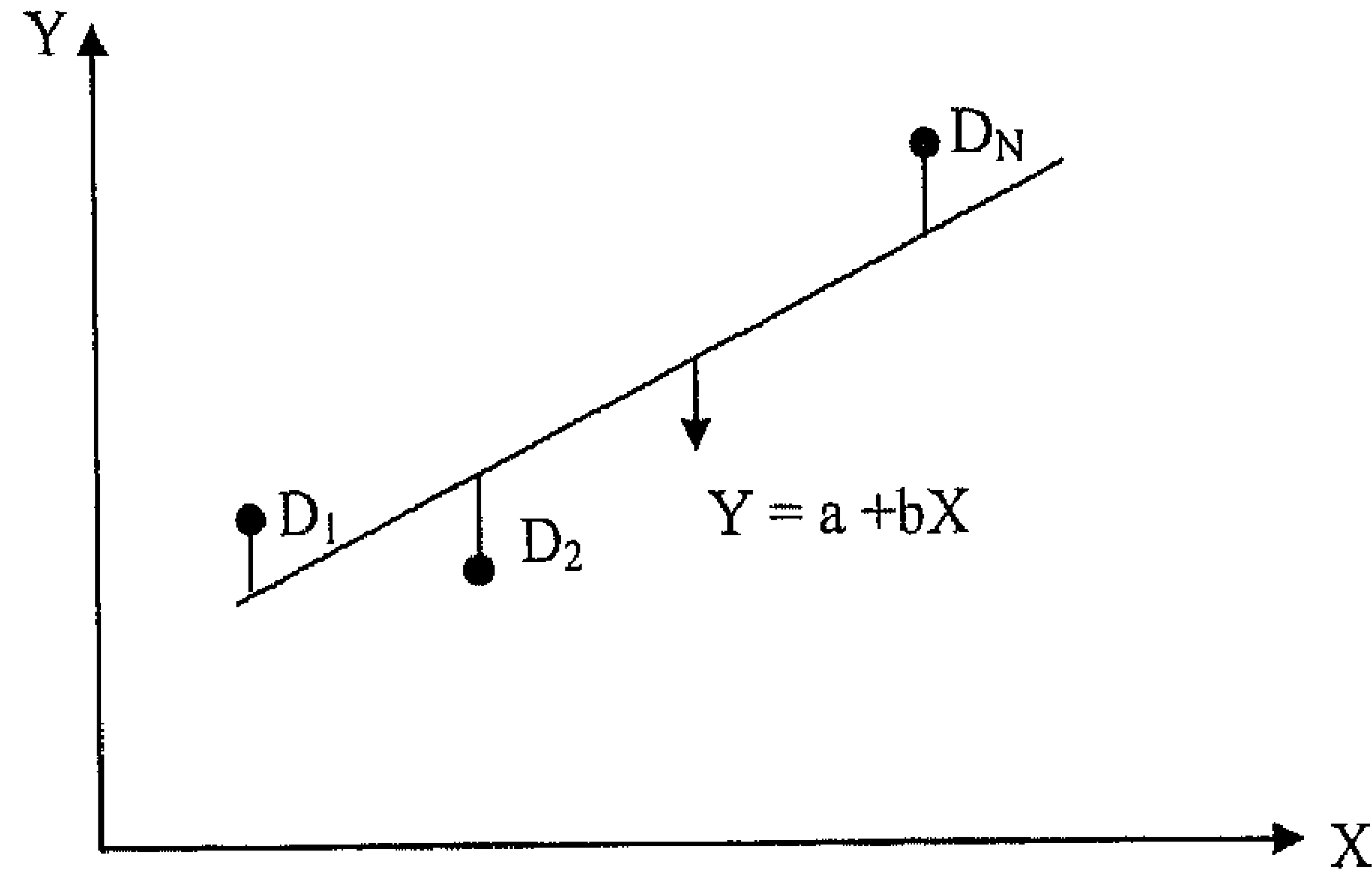
عدد الأولاد	1	2	3	4	5	المجموع
التكرار	5	6	5	5	4	25

وأيضا يمكن إنشاء الجدول التكراري المزدوج لمتغيرين X, Y إذا كان كل من المتغيرين من النوع الكمي المنفصل أو المتصل أو أحداهما كمياً منفصلاً والأخر كمياً متصلاً.

٥-٤ الانحدار الخطي البسيط (Simple Linear Regression)

من شكل الانتشار يمكن تحديد نوع العلاقة تقريباً بين المتغيرين وتختلف طريقه التقريب التي يمكن بها تعيين تلك العلاقة تبعاً للباحث والطريقة التي يمكن استخدامها للحصول على منحنى (علاقة رياضية) تقرب مجموعة النقاط بشكل الانتشار. ولذلك يتحتم علينا الالتزام بمقياس محدد للحكم على جودة التقريب بحيث يصبح الخطأ من عملية التقريب أقل ما يمكن. والطريقة المتبعة لتحقيق الغرض السابق تسمى بطريقة المربعات الصغرى (least square method) وتعتمد أساساً على وضع مقياس رياضي لمدى جودة توفيق منحنى معين لمجموعة من البيانات.

لنفرض أن لدينا مجموعة المشاهدات $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ وأنها قربت بالمنحنى $Y = a + bX$ وأن $D_1, D_2, \ldots, D_n$ تمثل الانحراف (الرأسي) بين البيانات الأصلية والمنحنى الذي يقربها كمقياس لجودة توفيق المنحنى للبيانات



وتعتبر الكمية $D_1^2 + D_2^2 + \cdots + D_n^2$ مجموع مربعات انحرافات البيانات عن المنحنى. ونلاحظ أنه إذا كان ذلك المجموع صغيراً فإن التوفيق (التقريب) يكون جيد وإذا كان كبيراً دل ذلك على رداءة المنحنى لتقريب للبيانات.

تعريف (٥-١)

من بين كل المنحنيات الممكنة كتقريب لمجموعة البيانات $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ فإن المنحنى الذي يتميز بأن قيمة $D_1^2 + D_2^2 + \cdots + D_n^2$ أقل ما يمكن يسمى أفضل منحنى تقريب لهذه البيانات. كما يسمى ذلك المنحنى بمنحنى المربعات الصغرى نسبة إلى الطريقة التي اتبعت في إيجاده.

وسندرس الآن بالتفصيل حالة شكل الانتشار والذي يمكن تقريبه بخط مستقيم (أي أن العلاقة بين المتغيرين X, Y هي علاقة خطية) واستخدام طريقة المربعات الصغرى للتعبير عن أفضل خط مستقيم يمكن أن يقرب البيانات المعطاة. وكيفية استخدام ذلك الخط في التنبؤ (الحصول على قيم تقديرية لبعض التجارب التي لم يتم إجرائها بعد). ويسمى الخط الذي يتم الحصول عليه بطريقة المربعات الصغرى بخط الانحدار لأحد المتغيرين على الآخر.

إذا كانت العلاقة بين المتغيرين X, Y يمكن تقريبها لخط مستقيم فتسمى العلاقة بين المتغيرين بعلاقة خطية ويمكن إيجاد أفضل خط مستقيم يستخدم لتعيين قيم Y بدلالة قيم X بالصورة $Y = a + bX$ ويقرب بمجموعة النقط $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ حيث a, b ثوابت يمكن الحصول على قيمها باستخدام طريقة المربعات الصغرى وبذلك يتحدد الخط المستقيم تماماً رياضياً (وبالتالي يمكن رسمه بيانياً).

نلاحظ أن الانحرافات(الرأسية) لمجموعة البيانات الأصلية عن الخط المستقيم $Y = a + bX$ هي

$$
\begin{aligned}
D_1 &= Y_1 - (a + bX_1)\\
D_2 &= Y_2 - (a + bX_2)\\
D_3 &= Y_3 - (a + bX_3)\\
&\vdots\\
D_n &= Y_n - (a + bX_n)
\end{aligned}
$$

ومجموع مربعات هذه الانحرافات

$$S = \sum_{i=1}^{n} D_i^2 = \sum_{i=1}^{n} (a + bX_i - Y_i)^2$$

نلاحظ أن S تعتمد في النهاية على a, b فقط، أي أنها دالة في a, b وتكون S أقل ما يمكن عندما يتحقق الشرطان

$$\frac{\partial S}{\partial a} = \frac{\partial}{\partial a} \sum_{i=1}^{n} (a + bX_i - Y_i)^2 = 0$$

$$\frac{\partial S}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} (a + bX_i - Y_i)^2 = 0$$

بإجراء التفاضل الجزئي بالنسبة إلى a والمساواة بالصفر. وكذلك بالنسبة إلى b والمساواة بالصفر يمكن كتابة الشرطين السابقين على الصورة التالية:

$$\sum_{i=1}^{n} Y_i = na + b \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = a \sum_{i=1}^{n} X_i + b \sum_{i=1}^{n} X_i^2$$

بحل المعادلتين الأخيرتين في a, b نحصل على

$$b = \frac{n\sum xy - (\sum X)(\sum y)}{n\sum x^2 - (\sum x)^2} = \frac{S_{xy}}{S_{xx}}$$

$$a = \bar{Y} - b\bar{X}$$

حيث:

$$S_{xy} = \sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X}) = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}$$

$$S_{xx} = \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

وبتعيين كل من a, b فإننا نحصل على أفضل خط مستقيم على الصورة

$$Y = a + bX$$

ونلاحظ من معادلة الخط المستقيم أن X تظهر كمتغير مستقل، Y متغير تابع ويسمى ذلك بخط انحدار Y على X. ويستخدم خط انحدار Y على X في التنبؤ عندما يكون لدينا قيمة معلومة للمتغير المستقل X ونريد إيجاد قيمة Y المناظرة. وأيضا يمكن تعيين خط انحدار X على Y والذي يستخدم لتعيين قيمة X إذا كانت قيمة Y معلومة.

مثال (٥-٥)

ارسم شكل الانتشار لمجموعة البيانات

X	2	4	1	3	5	6
Y	4	5	6	3	4	2

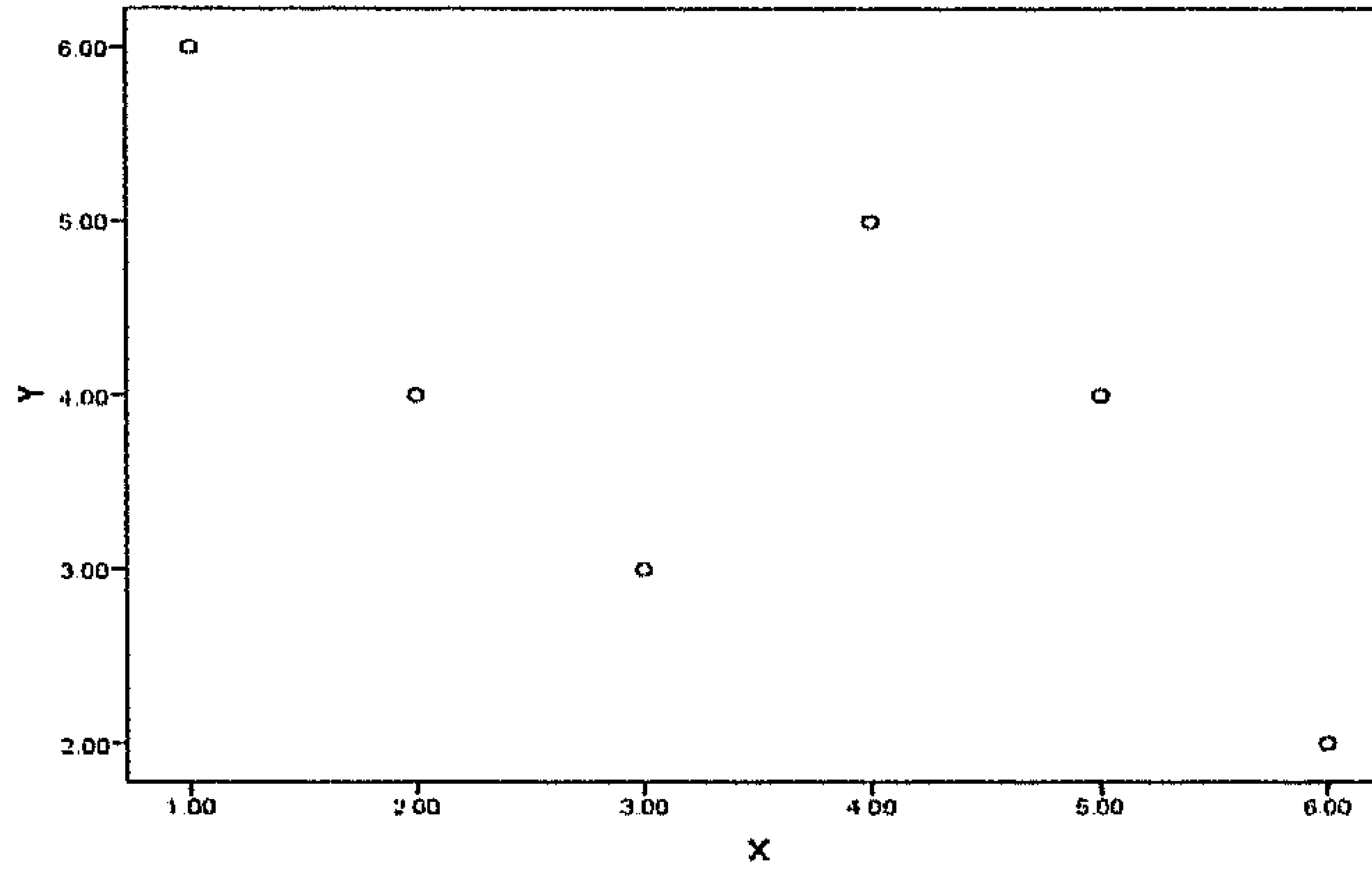
ثم أوجد

(أ) خط انحدار Y على X ومنه قدر قيمة Y عندما $X = 3.5$

(ب) خط انحدار X على Y ومنه قدر قيمة X عندما $Y = 2.3$

الحل

شكل الانتشار



من شكل الانتشار فإنه يمكن تقريب العلاقة بين المتغيرين X, Y كعلاقة خطية.

(أ) خط انحدار Y على X يمكن كتابته على الصورة

$$Y = a + bX$$

حيث a, b ثوابت.

ولحساب المعاملات العددية في المعادلتين الأخيرتين نكون جدولاً كالتالي

X	Y	XY	X^2
2	4	8	4
4	5	20	16
1	6	6	1
3	3	9	9
5	4	20	25
6	2	12	36
21	24	75	91

بالتعويض في المعادلات الاعتدالية لتعيين الثوابت a, b فإن

$$b = \frac{n\sum xy - (\sum X)(\sum y)}{n\sum x^2 - (\sum x)^2} = \frac{6(75) - (21)(24)}{6(91) - (21)^2} = -0.5143$$

$$a = \bar{Y} - b\bar{X} = \frac{24}{6} - \frac{(-0.5143)(21)}{6} = 5.8001$$

ويصبح خط الانحدار على الصورة التالية:

$$Y = 5.8001 - 0.5143X$$

عندما $X = 3.5$ فإن القيمة المقدرة للمتغير Y هي

$$Y = 5.8001 - 0.5143(3.5) = 4.0001$$

(ب) خط انحدار X على Y يمكن كتابته على الصورة

$$X = c + dY$$

حيث c, d ثوابت.

ولحساب المعاملات العددية في المعادلتين الأخيرتين نكون جدولاً كالتالي

X	Y	XY	Y^2
2	4	8	16
4	5	20	25
1	6	6	36
3	3	9	9
5	4	20	16
6	2	12	4
21	24	75	106

بالتعويض في المعادلات الاعتدالية لتصبح

$$d = \frac{n\sum_{i=1}^{n} X_iY_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n\sum_{i=1}^{n} {Y_i}^2 - (\sum_{i=1}^{n} Y_i)^2}$$

$$= \frac{6(75) - (21)(24)}{6(106) - (24)^2} = -0.9$$

$$c = \frac{\sum_{i=1}^{n} X_i}{n} - d\frac{\sum_{i=1}^{n} Y_i}{n} = \frac{21}{6} - \frac{(-0.9)(24)}{6} = 7.1$$

ويصبح خط الانحدار على الصورة التالية:

$$X = 7.1 - 0.9Y$$

عندما $Y = 2.3$ فإن القيمة المقدرة للمتغير X هي

$$X = 7.1 - 0.9(2.3) = 5.03$$

٥-٥ الارتباط (Correlation)

سوف نتعامل مع بعض المقاييس الإحصائية التي تستخدم لتحديد مقدار الارتباط(العلاقة) بين متغيرين (ظاهرتين) أو أكثر ومنها معامل الارتباط لبيرسون، معامل ارتباط الرتب لسبيرمان، ومعامل الاقتران والتوافق.

٥-٥-١ معامل الارتباط الخطي لبيرسون (Linear Correlation Coefficient of Pearson)

لدراسة الارتباط بين المتغيرين الكميين (X, Y) نحتاج لمقياس يحدد لنا مدى قوة العلاقة بين المتغيرين وأيضاً اتجاه هذه العلاقة، هل هي علاقة طردية بمعنى أن المتغيرين يتبعان نفس السلوك فزيادة قيم X تزيد قيم Y أو بنقصان قيم X تنقص قيم Y أم أنها علاقة عكسية بمعنى أن سلوك المتغيرين مختلف فزيادة قيم X تنقص قيم Y أو بنقصان قيم X تزيد قيم Y.

فإذا كان (X, Y) يعبران عن ظاهرتين مثل الطول (X) والوزن (Y) ومما سبق في الفصل الثالث فإن أفضل طريقة لمقارنة قيم ظاهرتين مختلفتين في وحدة القياس هو مقارنة القيم المعيارية لهما

$$Z_x = \frac{X - \bar{X}}{S_x}, \quad Z_y = \frac{Y - \bar{Y}}{S_y}$$

حيث S_X, S_Y هما الانحراف المعياري للمتغيرين X, Y وأيضا $\bar{X}, \bar{Y}$ هما المتوسط لقيم المتغيرين X, Y وقد قام بيرسون بأخذ متوسط حاصل ضرب القيمتين المعياريتين كمقياس لدرجة الارتباط بين المتغيرين (X, Y) وأطلق عليه معامل الارتباط ويرمز له بالرمز r حيث

$$r = \frac{1}{n}\sum_{i=1}^{n} Z_{x_i} Z_{y_i} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S_x}\right)\left(\frac{Y_i - \bar{Y}}{S_y}\right)$$

وبالتعويض عن قيم S_X, S_Y وأيضا $\bar{X}, \bar{Y}$ فإننا نحصل على الصيغة التالية لحساب معامل الارتباط الخطى لبيرسون

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2]\,[n\sum Y^2 - (\sum Y)^2]}}$$

$$= \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

حيث

$$S_{xy} = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}, \qquad S_{xx} = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

$$S_{yy} = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$

ويمكن إثبات أن قيمة معامل الارتباط تقع في الفترة $(-1,1)$ فإذا كانت إشارة الارتباط موجبة فإن الارتباط طردي لكن إذا كانت سالبة فإن الارتباط عكسي. ويمكن تقسيم الارتباط بناء على قيمته إلى:

(١) إذا كانت $r = 0$ فإنه لا يوجد ارتباط بين المتغيرين (X, Y)

(٢) إذا كانت $0 < r < 0.5$ فإن الارتباط ضعيف.

(٣) إذا كانت $r = 0.5$ فإن الارتباط متوسط.

(٤) إذا كانت $0.5 < r < 1$ فإن الارتباط قوى.

(٥) إذا كانت $r = 1$ فإن الارتباط تام.

ومعامل ارتباط بيرسون يستخدم في حالة البيانات الكمية ويشترط أن تكون:

(٦) البيانات مزدوجة

(٧) العلاقة بين المتغيرين علاقة خطية

(٨) بيانات المتغيرين تتبع التوزيع الطبيعي.

لذا يسمي معامل ارتباط بيرسون بمعامل الارتباط المعلمي.

مثال (٥-٦)

احسب معامل الارتباط الخطى لبيرسون بين الطول بالسم (Y) والوزن بالكيلوجرام (X) للبيانات التالية:

الوزن X	61	70	72	65	83	56	62
الطول Y	165	170	170	169	170	154	164

الحل:

معامل الارتباط الخطى لبيرسون يعطى من العلاقة التالية:

$$r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{(n\sum X^2 - (\sum X)^2)\,(n\sum Y^2 - (\sum Y)^2)}}$$

وبإنشاء الجدول التالي:

i	X	Y	XY	X^2	Y^2
1	61	165	10065	3721	27225
2	70	170	11900	4900	28900
3	72	170	12240	5184	28900
4	65	169	10985	4225	28561
5	83	170	14110	6889	28900
6	56	154	8624	3136	23716
7	62	164	10168	3844	26896
المجموع	469	1162	78092	31899	193098

ومن الجدول السابق فإن

$$r = \frac{7(78092) - (469)(1162)}{\sqrt{(7(31899) - (469)^2)\,(7(193098) - (1162)^2)}}$$

$$= \frac{1666}{2191.9726} = 0.76$$

ونجد أن الارتباط طردي قوى.

مثال (٥-٧)

في دراسة لمعرفة مدى العلاقة بين الدخل الشهري (X) والإنفاق الشهري (Y) قام أحد الباحثين بتجميع الدخل الشهري والإنفاق الشهري لثمانية أسر وكانت البيانات بالألف ريال سعودي كما يلي:

X	6	8	10	10	11	13	16	18
Y	6	7	10	8	9	11	16	17

والمطلوب تحديد نوع العلاقة بين الدخل الإنفاق الشهري.

الحل:

معامل الارتباط الخطى لبيرسون يعطى من العلاقة

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

بإنشاء الجدول التالي:

i	X	Y	XY	X^2	Y^2
1	6	6	36	36	36
2	8	7	56	64	49
3	10	10	100	100	100
4	10	8	80	100	64
5	11	9	99	121	81
6	13	11	143	169	121
7	16	16	256	256	256
8	18	17	306	324	289
المجموع	92	84	1076	1170	996

من الجدول السابق فإن

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{92}{8} = 11.5, \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{84}{8} = 10.5$$

وأيضاً

$$S_{xy} = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} = 1076 - 8(11.5)(10.5) = 110$$

$$S_{xx} = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = 1170 - 8(11.5)^2 = 112$$

$$S_{yy} = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2 = 996 - 8(10.5)^2 = 114$$

ومما سبق فإن معامل الارتباط الخطى لبيرسون

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{110}{\sqrt{112(114)}} = 0.973$$

ونجد أنه يوجد ارتباط طردي قوي بين الدخل الشهري والإنفاق الشهري.

٥-٥-٢ معامل ارتباط الرتب لسبيرمان (Rank Coefficient of Correlation for Spearman)

يستخدم معامل الارتباط للرتب لسبيرمان لدراسة الارتباط للبيانات الوصفية أو العددية، وتعتمد طريقة الارتباط للرتب على إعطاء المتغيرات رتباً تحل محل قيم المتغيرين (X, Y) ويرمز لمعامل الارتباط لسبيرمان بالرمز ρ والذي يعطى بالعلاقة التالية:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2-1)}$$

حيث n عدد أزواج القيم للمتغيرين (X,Y) وأن $\sum_{i=1}^{n} {d_i}^2$ مجموع مربعات الفروق بين رتبتي X, Y ويستخدم أحياناً معامل ارتباط سبيرمان للرتب مع البيانات الكمية. وسوف نقوم بترتيب قيم X تصاعدي/تنازلي معاً ثم نقوم بترتيب قيم Y تصاعدي/تنازلي وإذا وجدت قيم مكررة فإنه يتم إعطاء لكل منها رتبة تساوى متوسط رتب القيم المكررة وفى هذه الحالة تكون قيمة ρ هى قيمة تقريبية. وفي معامل ارتباط بيرسون إذا لم يتحقق أن البيانات العددية للمتغيرين لا تتبع التوزيع الطبيعي فإننا سوف نستخدم معامل ارتباط سبيرمان للرتب، والذي يطلق عليه معامل الارتباط اللامعلمي.

مثال (٥-٨)

تم تسجيل تقديرات عشرة طلاب في اختبار مادتي 1080 إحص، 1080 كيم من طلاب المستوي الثاني بكلية العلوم بجامعة الخرج. والمطلوب تعيين معامل الارتباط بين تقديرات الطلاب في المادتين

رقم الطالب	تقدير 1080 إحص	تقدير 1080كيم
1	ضعيف جدا	مقبول
2	مقبول	جيد
3	ممتاز	جيد جدا
4	مقبول	مقبول
5	ضعيف	جيد
6	جيد جدا	مقبول
7	جيد	ممتاز
8	ضعيف	ضعيف جدا
9	مقبول	ضعيف
10	مقبول	جيد جدا

الحل:

بما أن المتغيرين تقدير الطلاب في مقرر 1080 إحص وأيضا تقدير 1080 كيم هي متغيرات وصفية فإنه يتم استخدام معامل ارتباط سبيرمان للرتب لتعيين الارتباط بين التقدير في المقررين.

معامل ارتباط سبيرمان للرتب يعطى من العلاقة التالية:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2-1)}$$

بتعيين رتب التقدير في كل من المقررين:

رقم الطالب	تقدير 1080إحص	تقدير 1080 كيم	رتب تقدير 1080 إحص	رتب تقدير 1080 كيم	d	d^2
1	ضعيف جدا	مقبول	1	4	-3	9
2	مقبول	جيد	5.5	6.5	-1	1
3	ممتاز	جيد جدا	10	8.5	1.5	2.25
4	مقبول	مقبول	5.5	4	1.5	2.25
5	ضعيف	جيد	2.5	6.5	-4	16
6	جيد جدا	مقبول	9	4	5	25
7	جيد	ممتاز	8	10	-2	4
8	ضعيف	ضعيف جدا	2.5	1	1.5	2.25
9	مقبول	ضعيف	5.5	2	3.5	12.25
10	مقبول	جيد جدا	5.5	8.5	-3	9
المجموع						83

من الجدول السابق فإن:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2-1)}$$

$$= 1 - \frac{6(83)}{10(100-1)} = 0.497$$

فإن الارتباط طردي ضعيف.

مثال (٥-٩)

فيما يلي درجات 12 طالباً من طلاب كلية العلوم بجامعة الخرج، في اختبار مقرري 1080 ريض (X)، 1080 حيا (Y) والمطلوب تعيين معامل ارتباط سبيرمان للرتب بين درجات الطلاب في المادتين

X	16	12	10	13	9	10	11	13	14	17	19	20
Y	20	11	14	16	15	10	17	11	12	20	18	12

الحل

نقوم أولاً بتعيين رتب كلاً من درجات الطلاب في المقررين

رقم الطالب	X	Y	رتب X	رتب Y	d	d^2
1	16	20	9	11.5	-2.5	6.25
2	12	11	5	2.5	2.5	6.25
3	10	14	2.5	6	-3.5	12.25
4	13	16	6.5	8	-1.5	2.25
5	9	15	1	7	-6	36
6	10	10	2.5	1	1.5	2.25
7	11	17	4	9	-5	25
8	13	11	6.5	2.5	4	16
9	14	12	8	4.5	3.5	12.25
10	17	20	10	11.5	-1.5	2.25
11	19	18	11	10	1	1
12	20	12	12	4.5	7.5	56.25
المجموع						178

من الجدول السابق فإن

$$\rho = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2-1)}$$
$$= 1 - \frac{6(178)}{12(144-1)} = 0.3776$$

وبالتالي فإن الارتباط طردي ضعيف.

٥-٥-٣ معامل الاقتران ومعامل التوافق

(Coefficient of Contingency and Coefficient of Association)

يستخدم معامل ارتباط بيرسون لتعيين قوة ونوع العلاقة في حالة البيانات الكمية ومعامل ارتباط سبيرمان للرتب يستخدم لتعيين قوة الارتباط للرتب مع كل من البيانات الكمية والنوعية التي لها صفة الترتيب، لكن هناك بيانات وصفية لها صفات مميزة لا نستطيع ترتيبها، مثل الحالة الاجتماعية (أعزب — متزوج - أرمل — مطلق) وكذلك عادة التدخين (مدخن — غير مدخن) وغير ذلك.

ولقياس قوة الارتباط لهذه البيانات نشأت الحاجة إلى مقاييس الارتباط بين الصفات. وسوف نهتم هنا بمعامل الاقتران لكرامير ويستخدم لتعيين قوة الارتباط بين ظاهرتين لكل منهما صفتان فقط، وكذلك دراسة معامل التوافق لكرامير ويستخدم إذا كان لكل ظاهرة أكثر من صفتين.

أولا معامل الاقتران (Coefficient of Contingency)

يستخدم معامل الاقتران لقياس قوة الارتباط بين ظاهرتين لكل منهما صفتان فقط، ويرمز لمعامل الاقتران بالرمز r_c وبفرض وجود ظاهرة X وظاهرة أخرى Y لكل منهما صفتان كما في الجدول التالي:

الظاهرة X / الظاهرة Y	الصفة الأولى	الصفة الثانية
الصفة الأولى	a	b
الصفة الثانية	c	d

فإن معامل الاقتران يمكن تعيينه باستخدام العلاقة التالية:

$$r_c = \frac{ad - cb}{ad + cb}$$

مثال (٥-١٠)

احسب معامل الاقتران بين التعليم والعمل لمجموعة من الأفراد حيث كانت بياناتهم كالتالي:

التعليم X / العمل Y	متعلم	غير متعلم
يعمل	12	7
لا يعمل	6	8

الحل

معامل الاقتران بين التعليم والعمل هو

$$r_c = \frac{ad - cb}{ad + cb} = \frac{12(8) - 6(7)}{12(8) + 6(7)} = 0.391$$

ونلاحظ أن الارتباط ضعيف.

ثانيا معامل التوافق (Coefficient of Association)

إذا كان لدينا ظاهرتان X, Y لكل منهما أكثر من صفتين فإن معامل الاقتران لا يصلح في هذه الحالة لذا سوف نستخدم مقياساً أخر وهو معامل التوافق والذي يرمز له بالرمز r_a ولتعيين معامل التوافق نفرض أن الظاهرة X لها r من الصفات وأيضا الظاهرة Y لها s من الصفات كما بالجدول المزدوج التالي بحيث $r, s > 2$

الظاهرة X / الظاهرة Y	Y_1	Y_2	...	Y_s	المجموع
X_1	f_{11}	f_{12}	...	f_{1s}	f_{1*}
X_2	f_{21}	f_{22}	...	f_{2s}	f_{2*}
...	...	...	...	...	...
X_r	f_{r1}	f_{r2}	...	f_{rs}	f_{r*}
المجموع	f_{*1}	f_{*2}	...	f_{*s}	f_{**}

حيث $j = 1, 2, \ldots, s$ $,i = 1, 2, \ldots, r,$ f_{ij} هى التكرارت المشاهدة للصفتين وايضاً

$$f_{i*} = \sum_{j=1}^{s} f_{ij}, \qquad f_{*j} = \sum_{i=1}^{r} f_{ij}$$

نقوم بحساب معامل التوافق من العلاقة التالية:

$$r_a = \sqrt{\frac{B-1}{B}}$$

حيث المقدار B يتم حسابه من العلاقة:

$$B = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(f_{ij})^2}{f_{i*} \times f_{*j}}$$

مثال (٥-١١)

احسب معامل التوافق بين لون الشعر ولون العيون لعينة من 45 فرداً والتي لخصت بياناتهم كما في الجدول المزدوج التالي:

لون الشعر / لون العين	أشقر	بنى	أسود	المجموع
أزرق	6	5	4	15
عسلي	3	6	6	15
أسود	2	7	6	15
المجموع	11	18	16	45

الحل

أولاً: نقوم بحساب قيمة B كما يلي:

$$B = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(f_{ij})^2}{f_{i*}\times f_{*j}}$$

$$= \frac{(6)^2}{15\times 11}+\frac{(5)^2}{15\times 18}+\frac{(4)^2}{15\times 16}+\frac{(3)^2}{15\times 11}+\frac{(6)^2}{15\times 18}$$

$$+\frac{(6)^2}{15\times 16}+\frac{(2)^2}{15\times 11}+\frac{(7)^2}{15\times 18}+\frac{(6)^2}{15\times 16} = 1.071$$

ثانياً: نقوم بتعيين قيمة معامل التوافق كما يلي

$$r_a = \sqrt{\frac{B-1}{B}} = \sqrt{\frac{1.071-1}{1.071}} = 0.257$$

ونلاحظ أن قيمة معامل التوافق ضعيفة، أي أن قوة الارتباط بين لون البشرة ولون العيون ضعيفة.

٥-٦ العلاقة بين معامل الارتباط لبيرسون ومعامل الانحدار b

لقد تم تعريف معامل الارتباط لبيرسون بالعلاقة التالية:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

وأيضاً تم تعريف معامل الانحدار الخطى البسيط بالعلاقة التالية:

$$b = \frac{S_{xy}}{S_{xx}} \Longrightarrow S_{xy} = b\,S_{xx}$$

وبالتعويض عن S_{xy} في معامل الارتباط نحصل على

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = b\frac{S_{xx}}{\sqrt{S_{xx}S_{yy}}} = b\sqrt{\frac{S_{xx}}{S_{yy}}} = b\frac{S_x}{S_y}$$

حيث

$$S_x = \sqrt{\frac{1}{n-1}\left[\sum X^2 - n\bar{X}^2\right]}, \qquad S_y = \sqrt{\frac{1}{n-1}\left[\sum Y^2 - n\bar{Y}^2\right]}$$

مثال (٥-١٢)

إذا علمت أن الوسط الحسابي والانحراف المعياري للمتغير X هما $\bar{X} = 4, S_x = 0.25$ وأن الوسط الحسابي والانحراف المعياري للمتغير Y هما $\bar{Y} = 6,\ S_y = 0.8$ فأوجد معادلة انحدار Y على X علماً بأن معامل الارتباط بين X, Y هو $r = 0.78$

الحل:

من العلاقة بين معامل الارتباط ومعامل الانحدار فإن

$$r = b\frac{S_x}{S_y} \Longrightarrow b = r\frac{S_y}{S_x} = 0.78\left(\frac{0.8}{0.25}\right) = 2.496$$

ومن المعلوم أن

$$a = \bar{Y} - b\bar{X} = 6 - (2.496)(4) = -3.984$$

وبذلك فإن خط انحدار Y على X هو

$$Y = a + bX = -3.984 + 2.496X$$

٥-٧ تطبيقات باستخدام برنامج SPSS

تطبيق (٥-١)

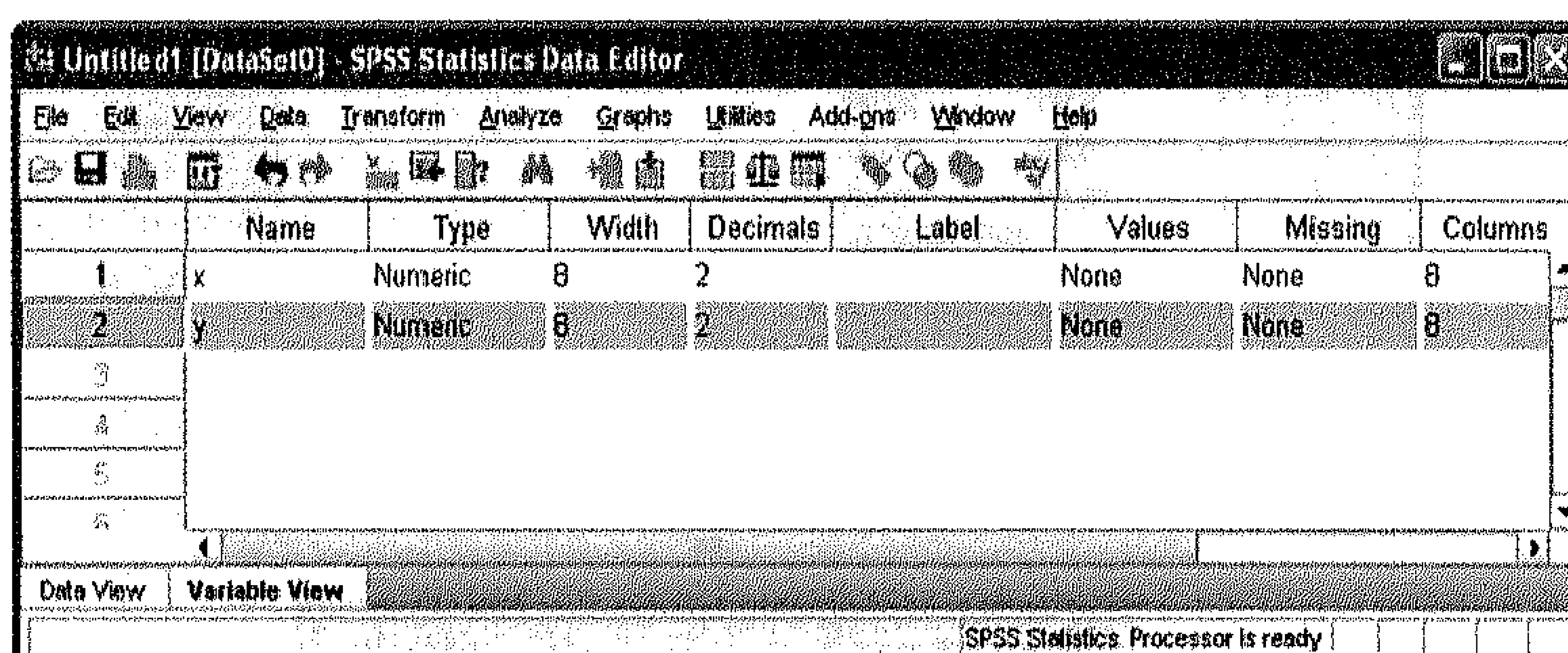
في مثال (٥-٢)، ارسم شكل الانتشار لمجموعة البيانات التالية وحدد نوع العلاقة بينهما

X	2	4	1	3	5	6	8	10
Y	7	13	3	11	17	20	27	33

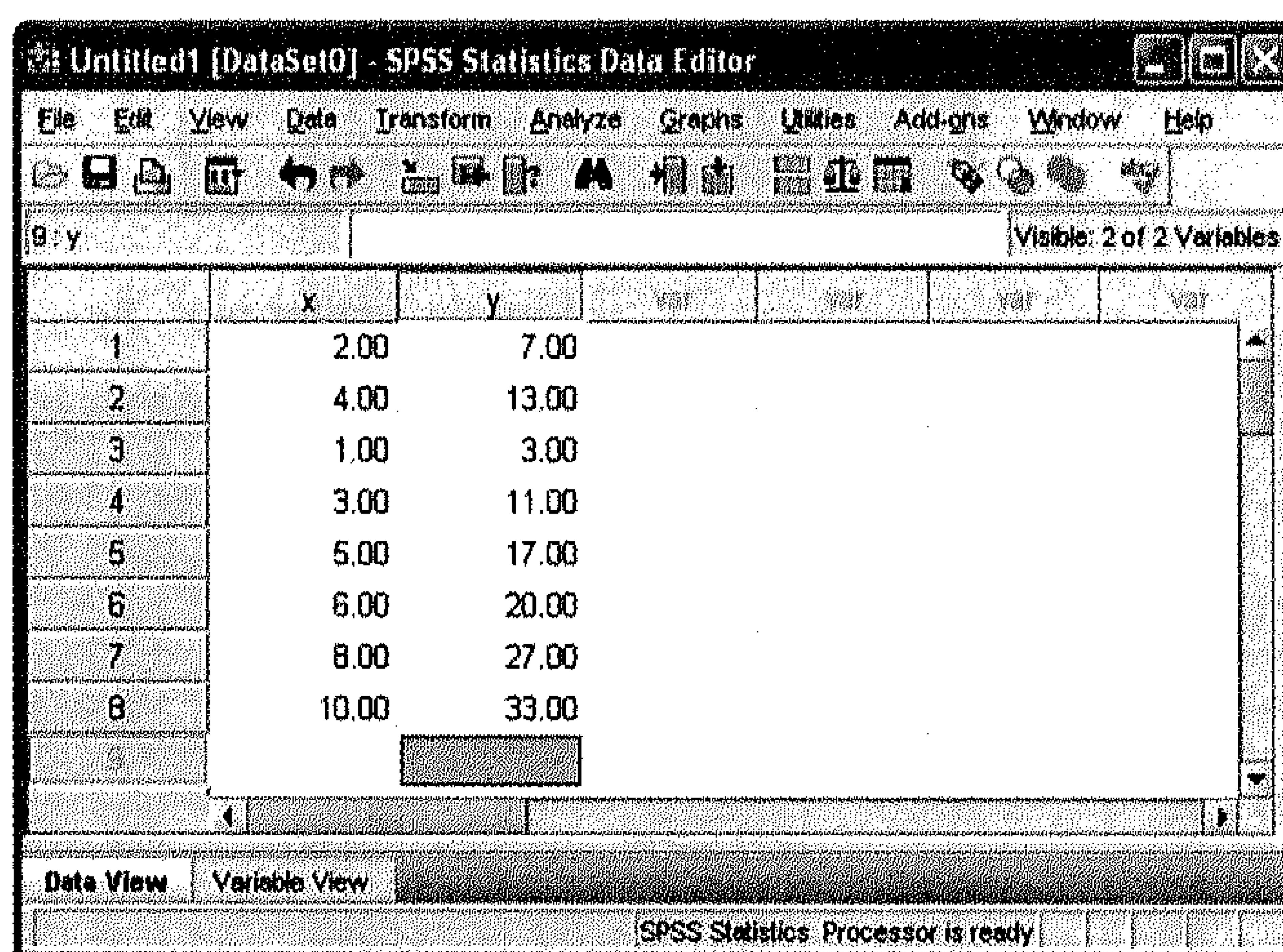
الحل

أولاً: نقوم بإدخال البيانات للبرنامج باتباع الخطوات التالية:

١- نقوم بالانتقال لنافذة Variable View ونقوم بتعريف المتغيرين X, Y

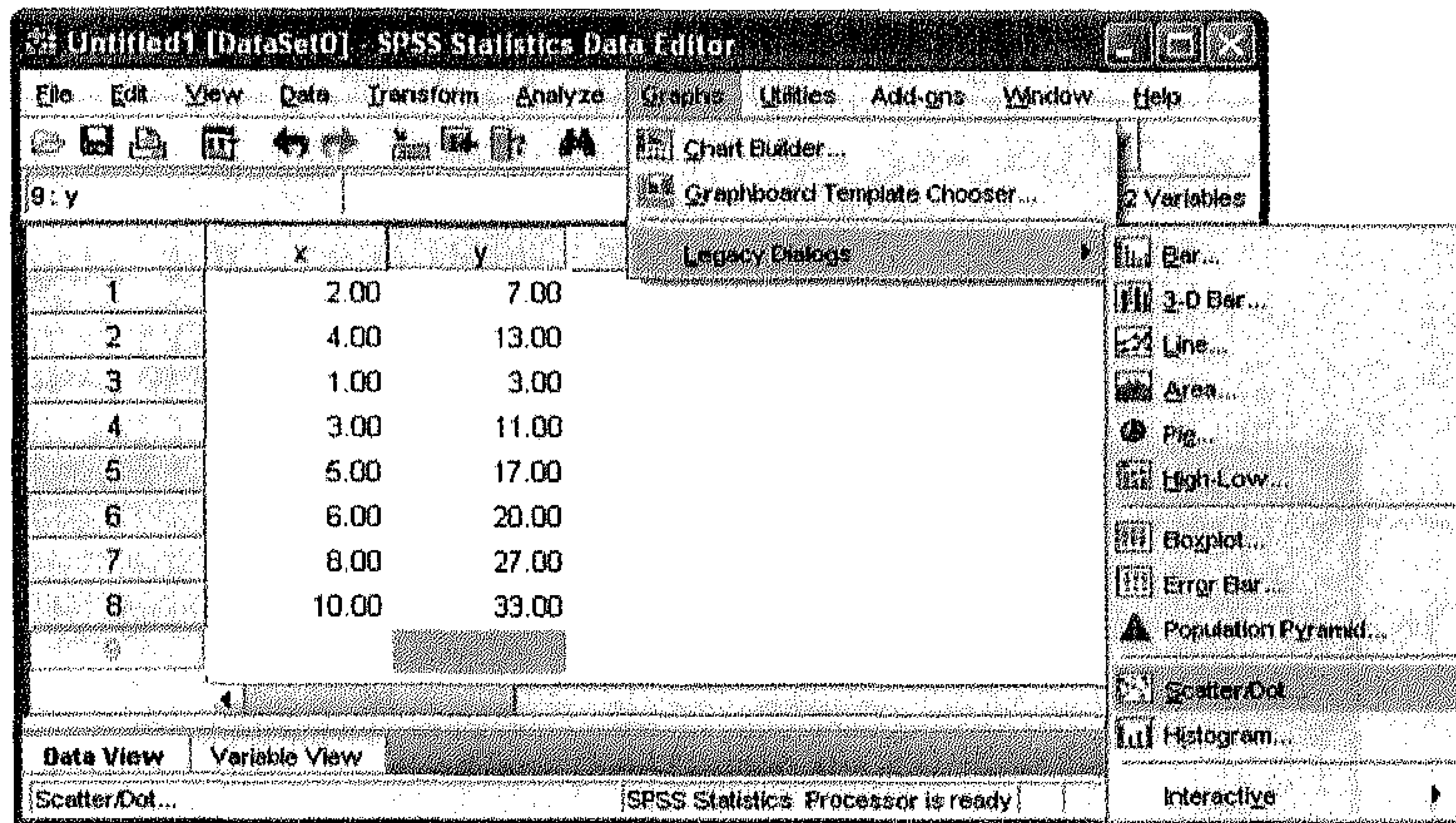


٢- ننتقل لنافذة Data View ونقوم بإدخال البيانات الخاصة بكل متغير

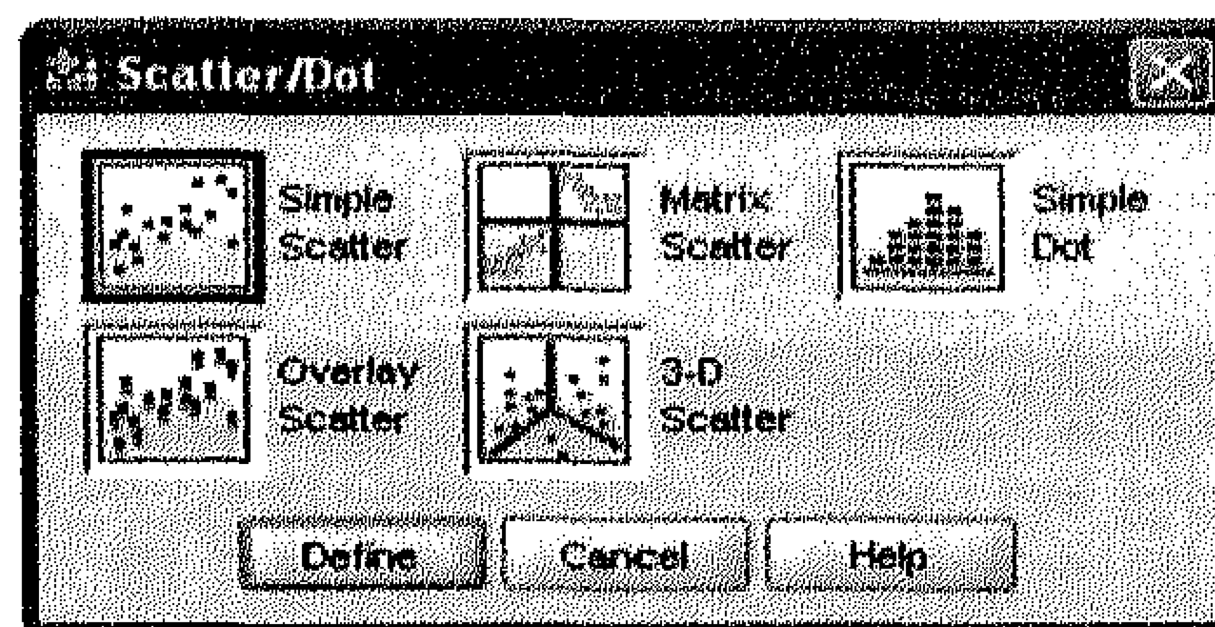


ثانياً: نقوم برسم شكل الانتشار بين المتغيرين X, Y باتباع الخطوات التالية:

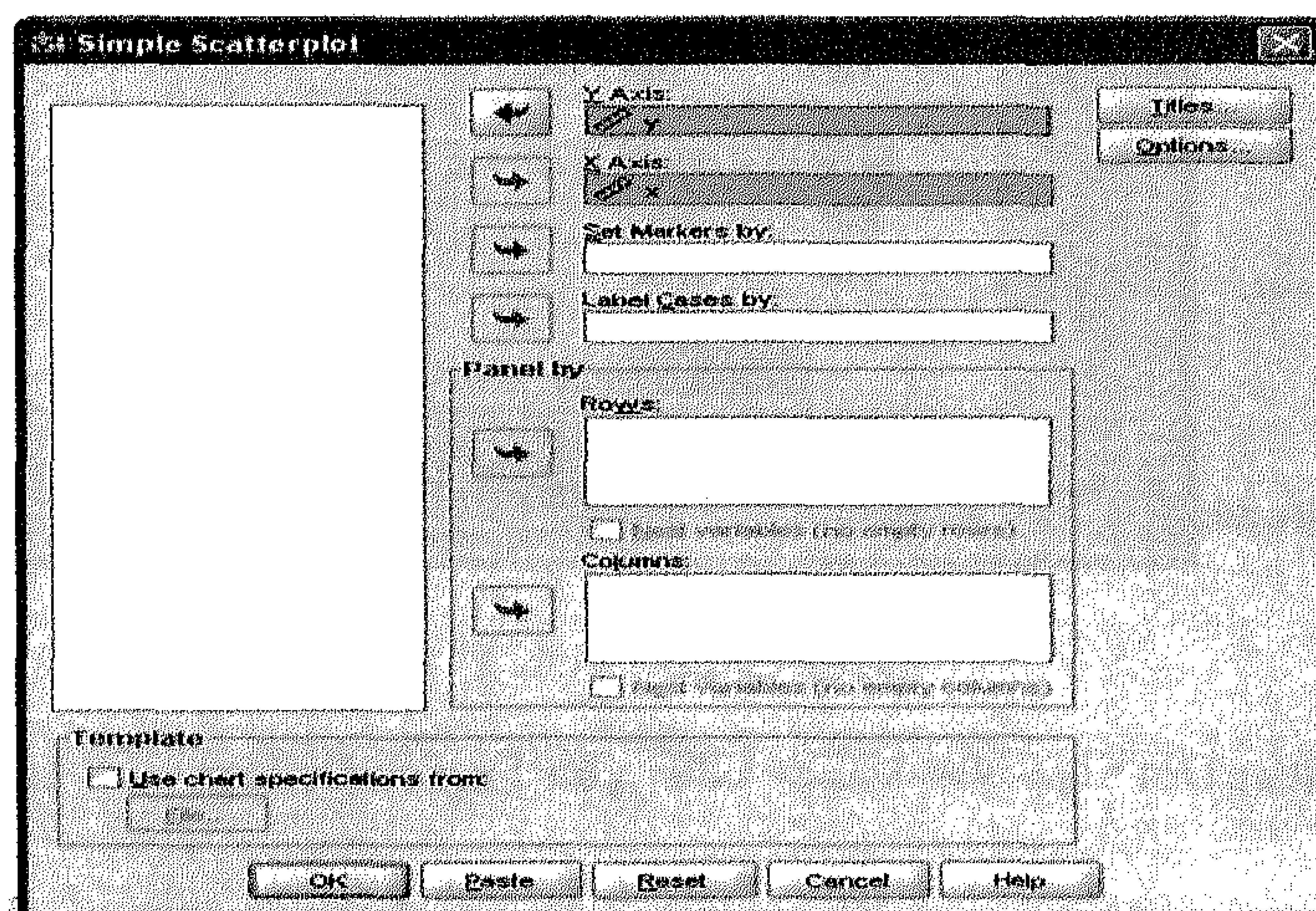
١- من قائمة Graphs نختار Legacy Dialogs فتظهر قائمة منسدلة نختار منها Scatter/Dot



٢- تظهر شاشة جديدة بعنوان Scatter/Dot فنختار منها Simple Scatter ثم نضغط على Define

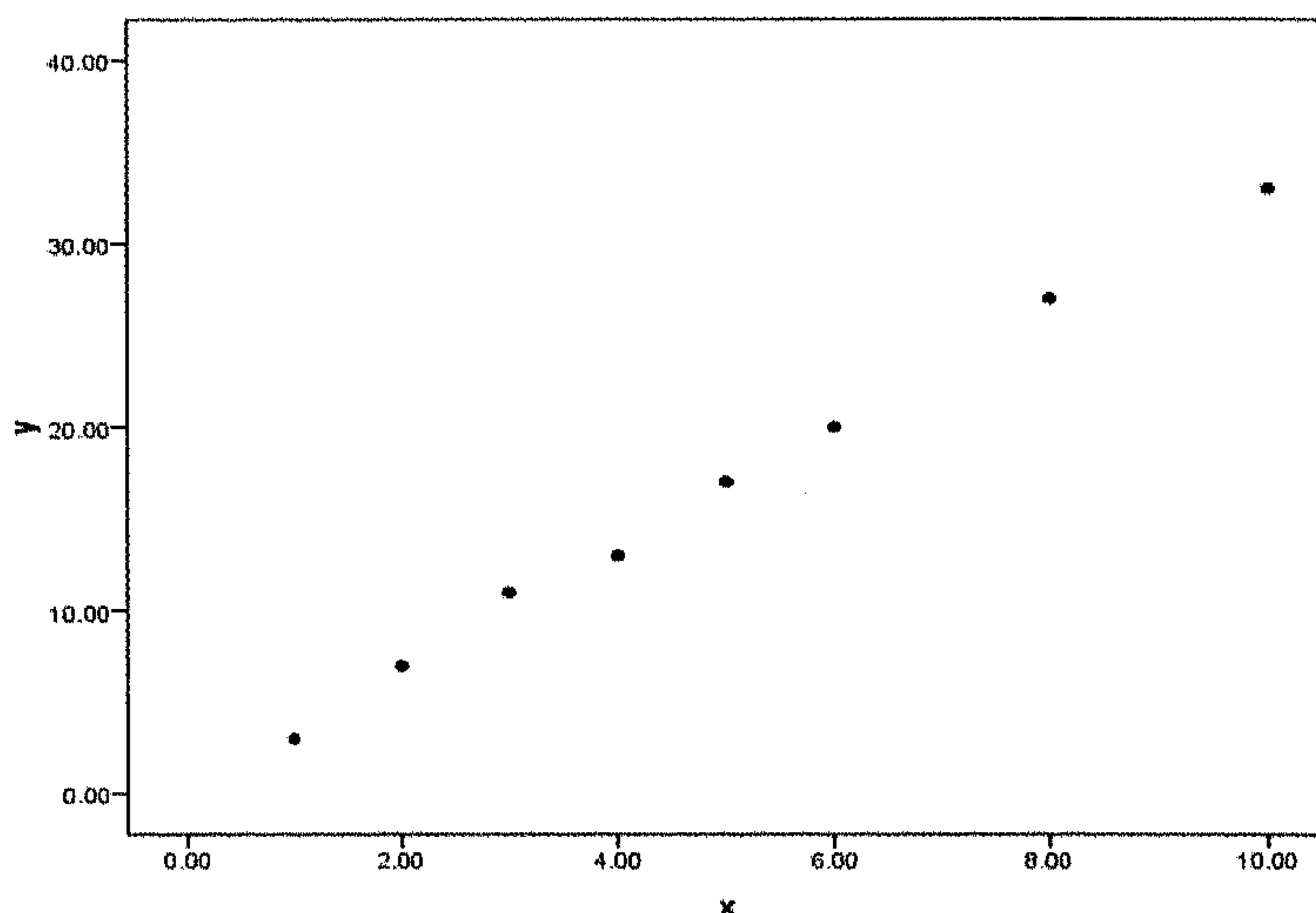


٣- تظهر نافذة جديدة بعنوان Simple Scatter plot



٤- ننقل المتغير X لخانة :X Axis وأيضا ننقل المتغير Y لخانة :Y Axisثم نضغط على Ok

٥- فننتقل لنافذة المخرجات Output View ويظهر الشكل البياني التالي الذي يمثل شكل الانتشار



ومن الشكل السابق نجد أن النقاط يمكن تقريبها بخط مستقيم، فإن العلاقة بين X,Y علاقة خطية طردية.

تطبيق (٥-٢)

في مثال (٥-٤): أراد أحد الباحثين دراسة العلاقة بين مستوى التعليم وعدد الأولاد فقام بأخذ عينة من 25 شخص من أحد الأحياء وقام بتوجيه بعض الأسئلة التي تتضمن مستوى التعليم للشخص وعدد الأولاد وكانت النتائج كالتالي:

مستوى التعليم X	ضعيف	متوسط	عالي	عالي	متوسط	ضعيف	ضعيف	متوسط	عالي
عدد الأولاد Y	4	3	5	4	2	4	3	2	5
مستوى التعليم X	متوسط	ضعيف	متوسط	عالي	عالي	متوسط	ضعيف	ضعيف	متوسط
عدد الأولاد Y	2	4	5	1	2	3	1	2	1
مستوى التعليم X	عالي	متوسط	ضعيف	متوسط	متوسط	عالي	ضعيف		
عدد الأولاد Y	1	2	3	4	5	1	3		

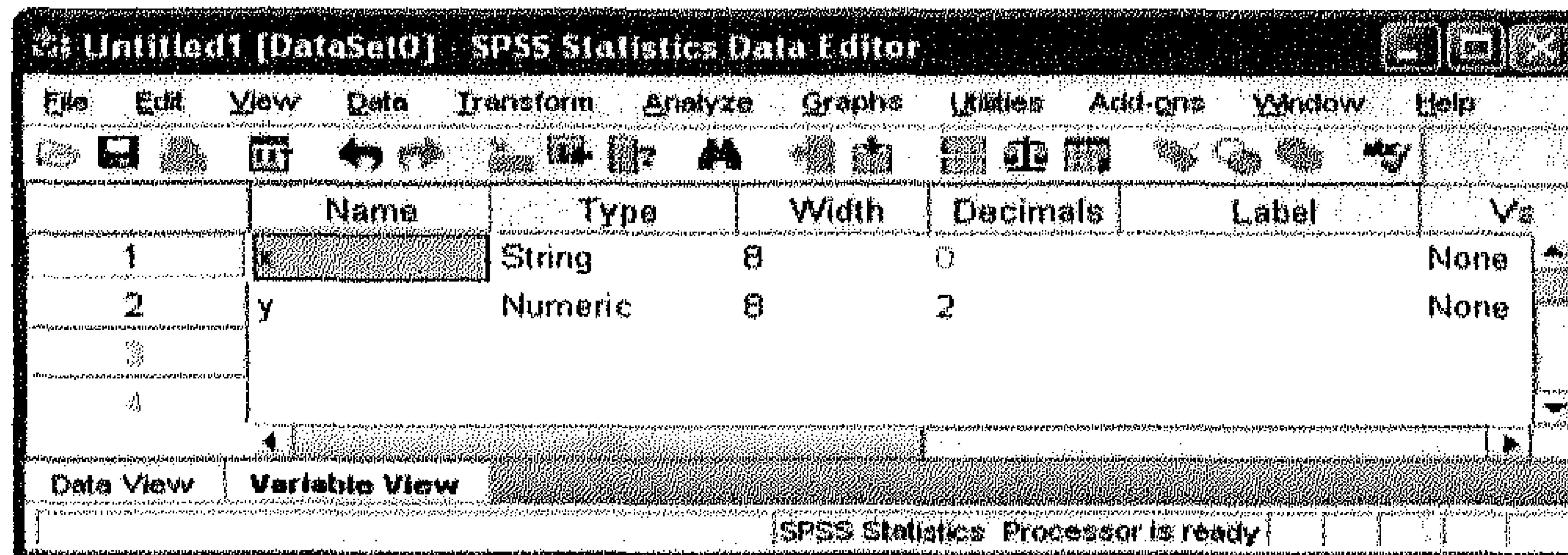
كون الجدول التكراري المزدوج لمستوى التعليم (X) وعدد الأولاد (Y) .

الحل

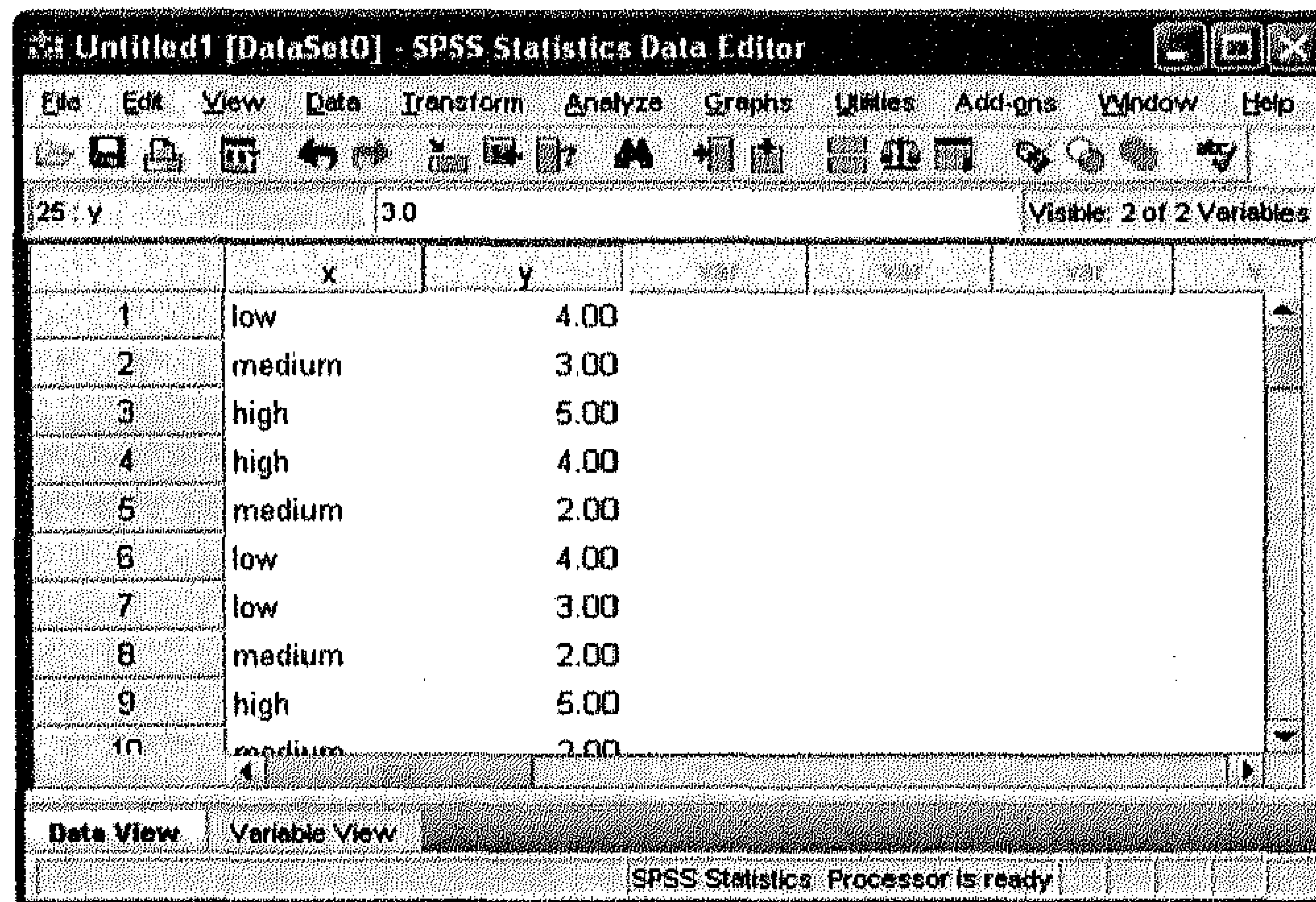
الإصدار 17 من برنامج SPSS لا يقبل التعامل مع البيانات المكتوبة باللغة العربية لذا سوف نترجم الكلمات العربية للغة الانجليزية ونتعامل معها.

أولاً: بإدخال قيم المتغيرين X, Y باتباع الخطوات التالية:

١- نقوم بتعريف المتغيرين X, Y في نافذة Variable View

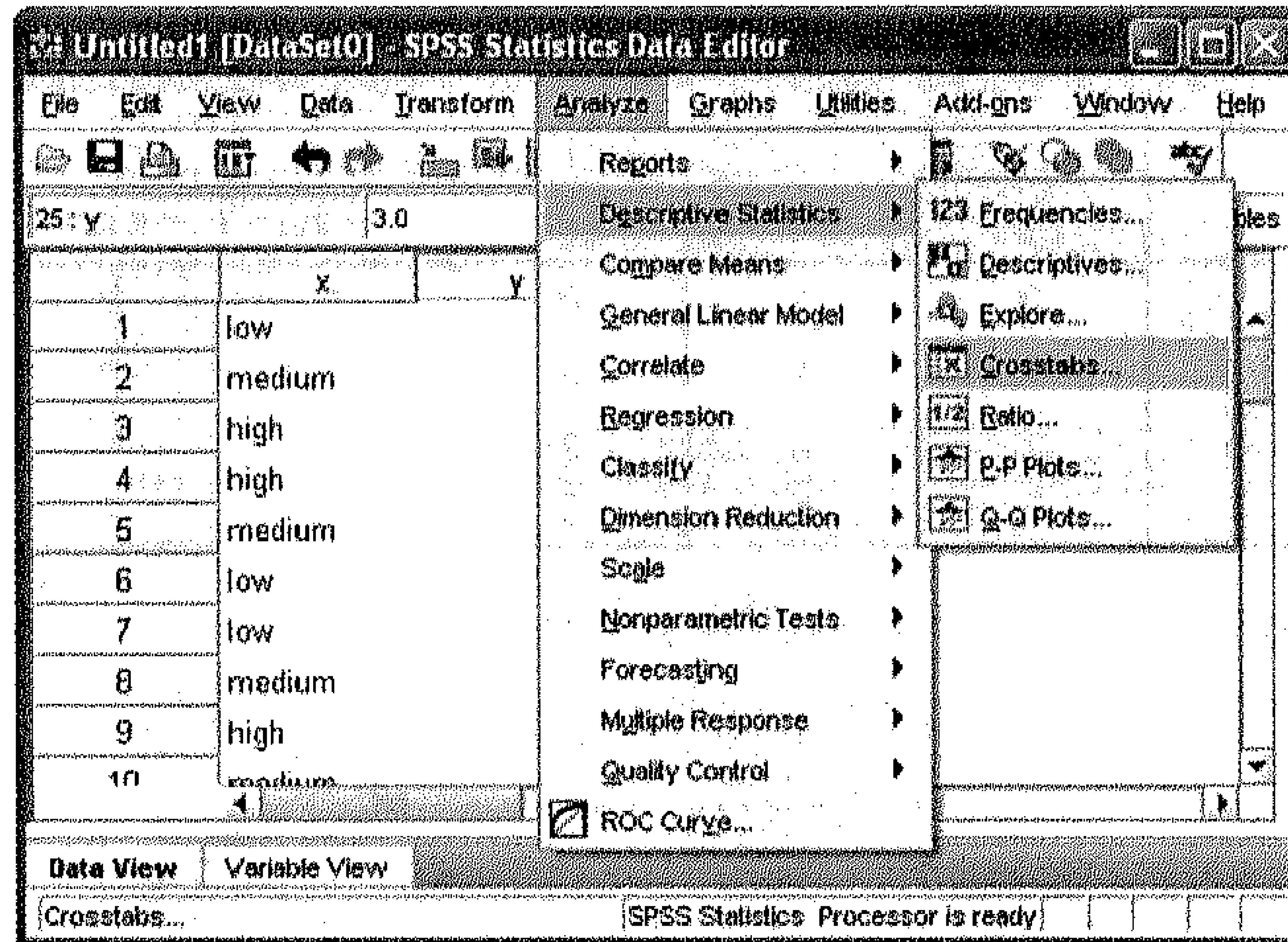


٢- نتنقل لنافذة Data View ونقوم بإدخال البيانات الخاصة بكل متغير

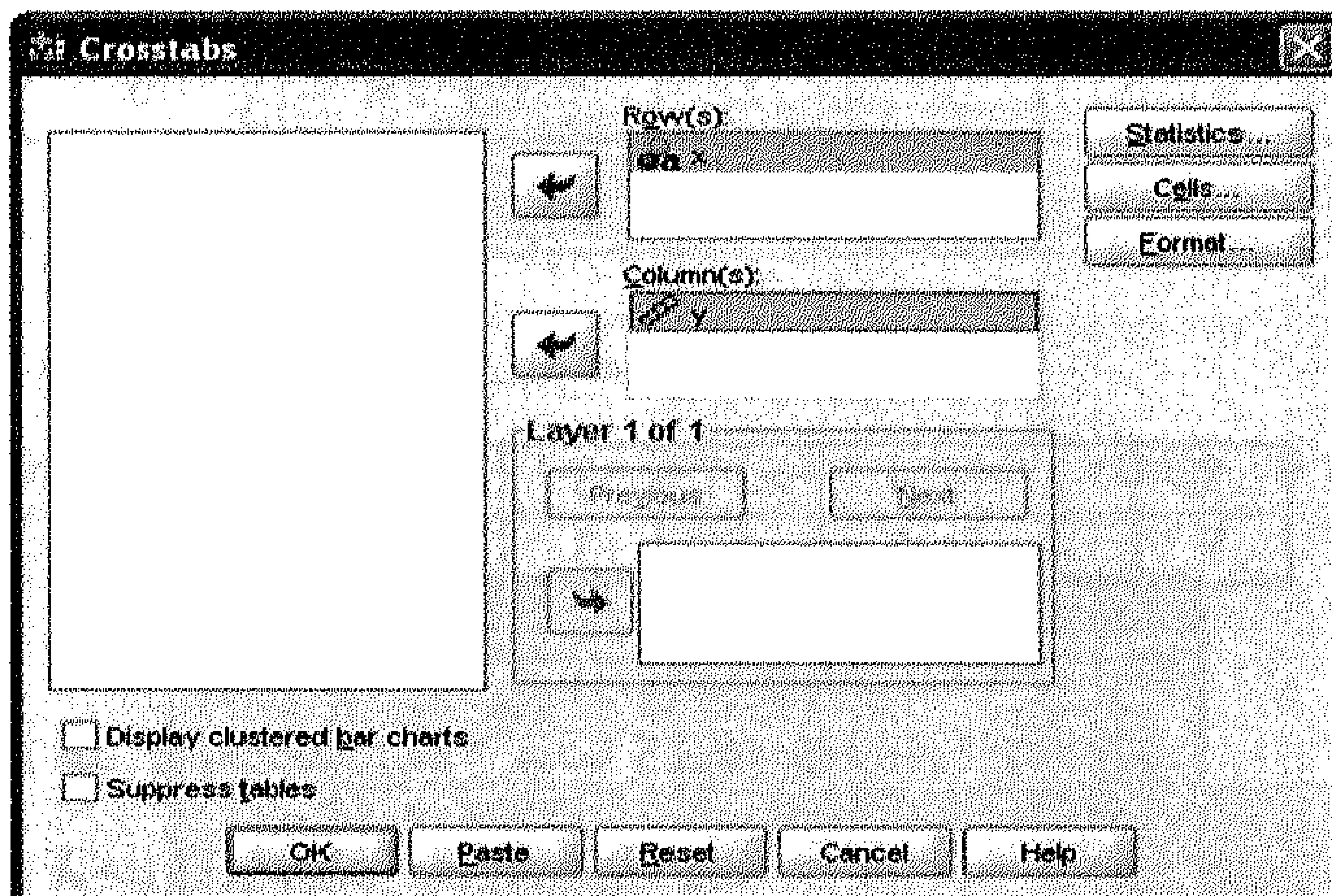


ثانياً: لتكوين الجدول التكراري المزدوج للمتغيرين X، Y نتبع الخطوات التالية:

١- من قائمة Analyze نختار Descriptives Statistics



٢- من القائمة المنسدلة نختار Crosstabs فتظهر شاشة جديدة بعنوان Crosstabs ننقل المتغير X لقائمة Row(s) وننقل المتغير Y لقائمة Column(s)



٣- نضغط على Ok تظهر النتائج التالية

x * y Crosstabulation

Count

		y					
		1.00	2.00	3.00	4.00	5.00	Total
x	high	3	1	0	1	2	7
	low	1	1	3	3	0	8
	medium	1	4	2	1	2	10
Total		5	6	5	5	4	25

ومن الجدول السابق يمكن تعيين التوزيع التكراري الهامشي لكل من X وأيضا Y حيث التوزيع التكراري الهامشي للمتغير X هو

X	ضعيف	متوسط	عالي	المجموع
التكرار	8	10	7	15

التوزيع التكراري الهامشي للمتغير Y هو

Y	1	2	3	4	5	المجموع
التكرار	5	6	5	5	4	15

تطبيق (٥-٣)

في مثال (٥-٥): ارسم شكل الانتشار لمجموعة البيانات

X	2	4	1	3	5	6
Y	4	5	6	3	4	2

ثم أوجد

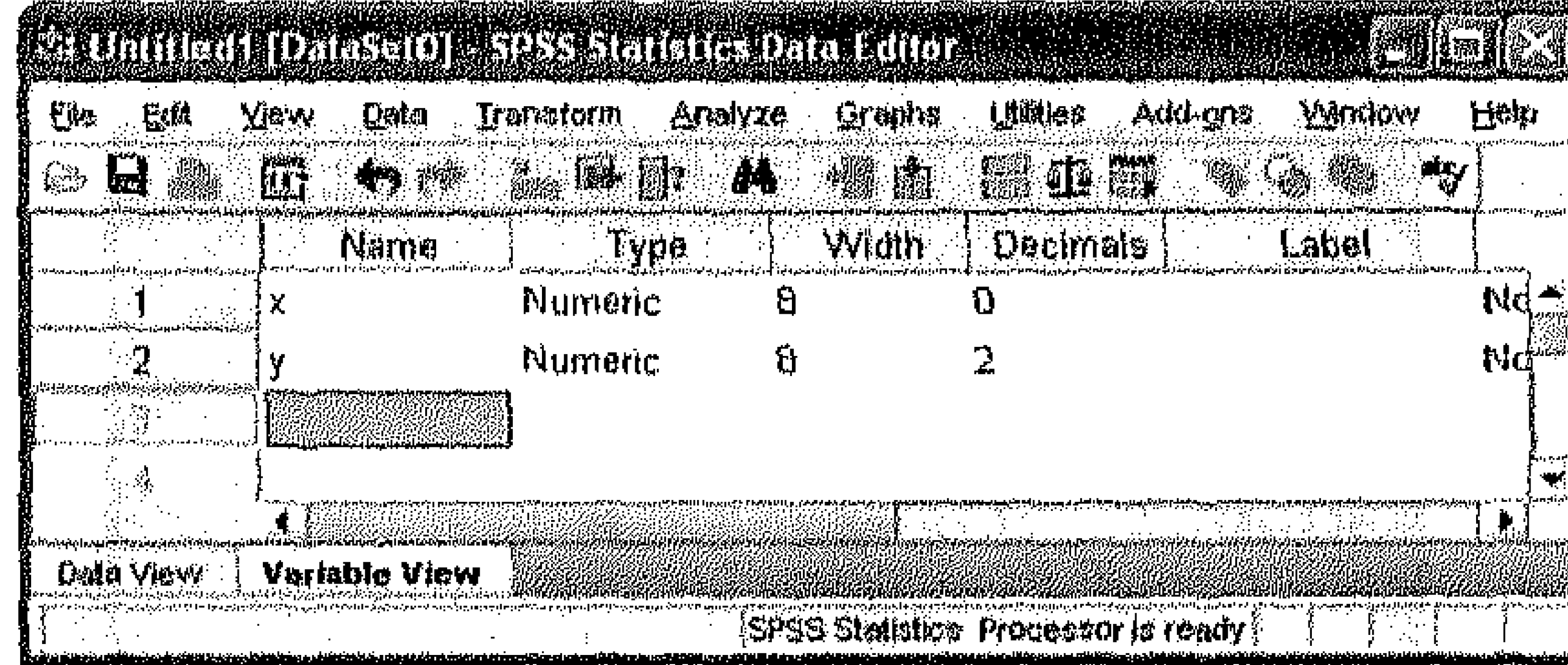
(أ) خط انحدار Y على X ومنه قدر قيمة Y عندما $X = 3.5$

(ب) خط انحدار X على Y ومنه قدر قيمة X عندما $Y = 2.3$

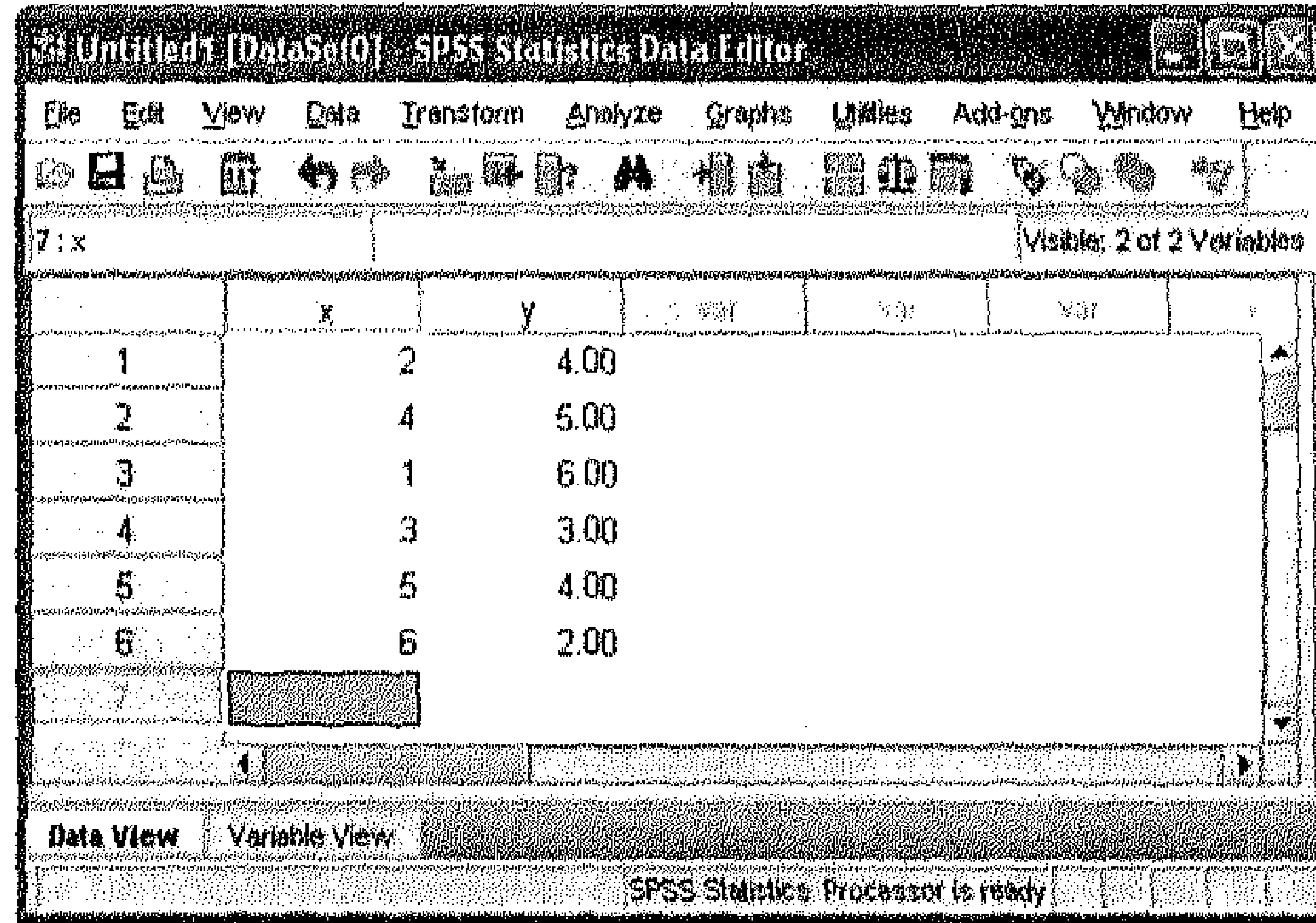
الحل

أولاً: إدخال البيانات

١- سوف نقوم بتعريف المتغيرين X, Y باستخدام لنافذة Variable View

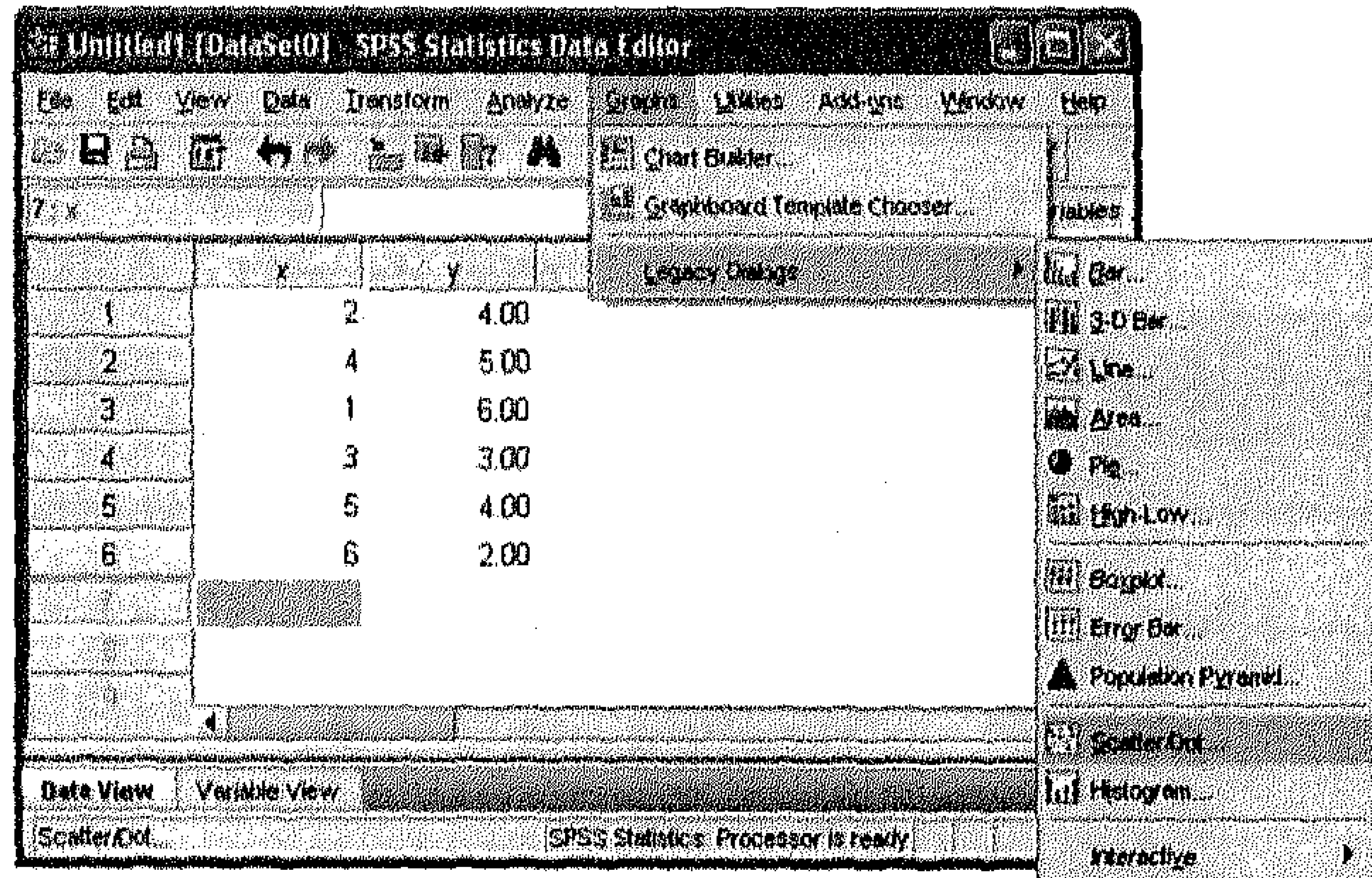


٢- ننتقل لنافذة Data View ونقوم بإدخال البيانات

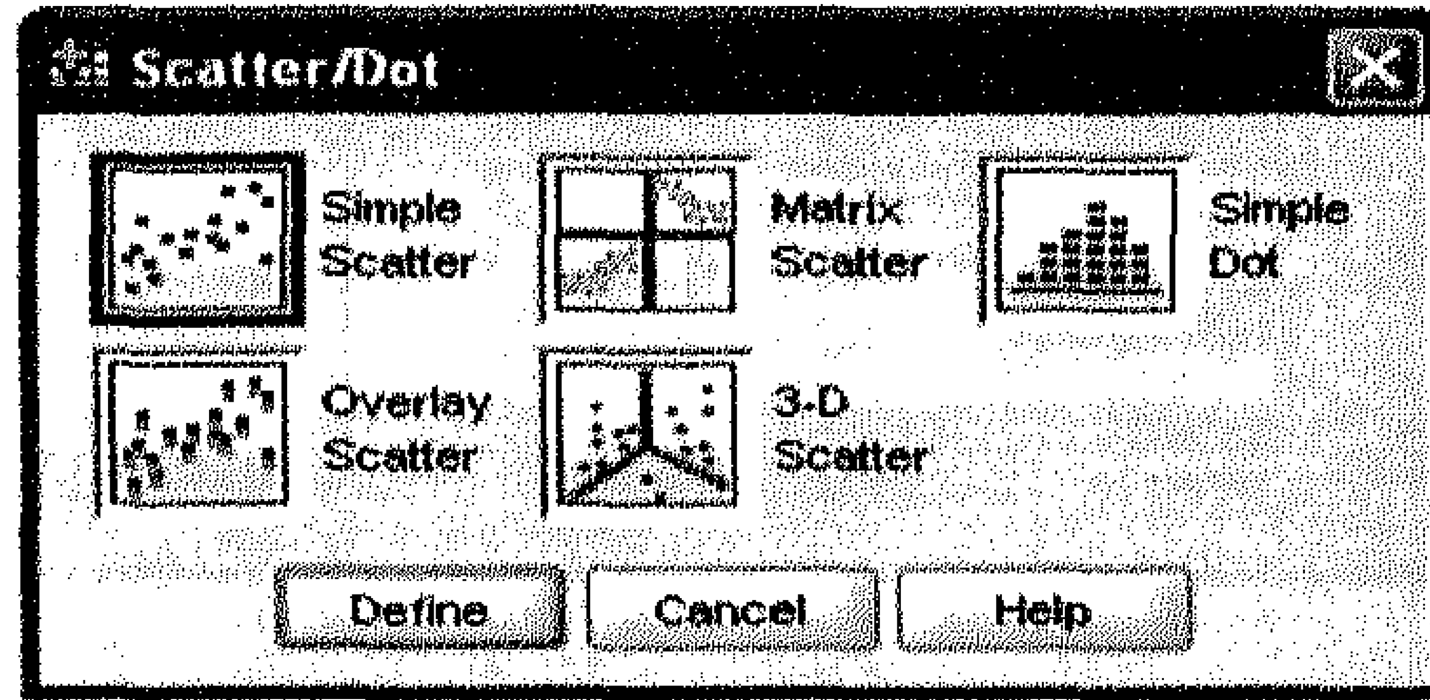


ثانياً: رسم شكل الانتشار

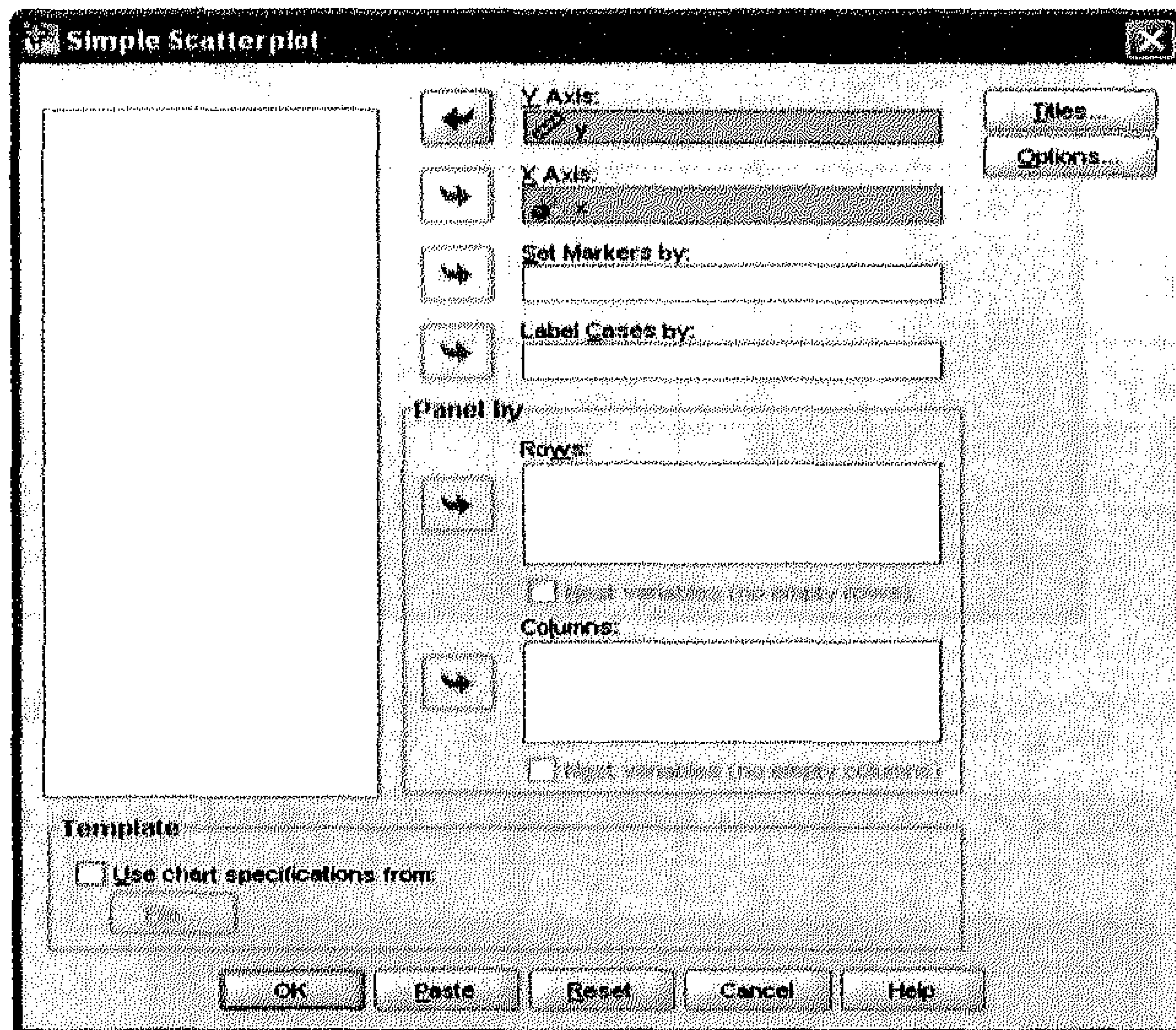
١- من قائمة Graphs نختار Legacy Dialogs فتظهر قائمة منسدلة فنختار منها Scatter/Dot



٢- فتظهر شاشة جديدة بعنوان Scatter/Dot نختار منها Simple Scatter ثم نضغط على Define

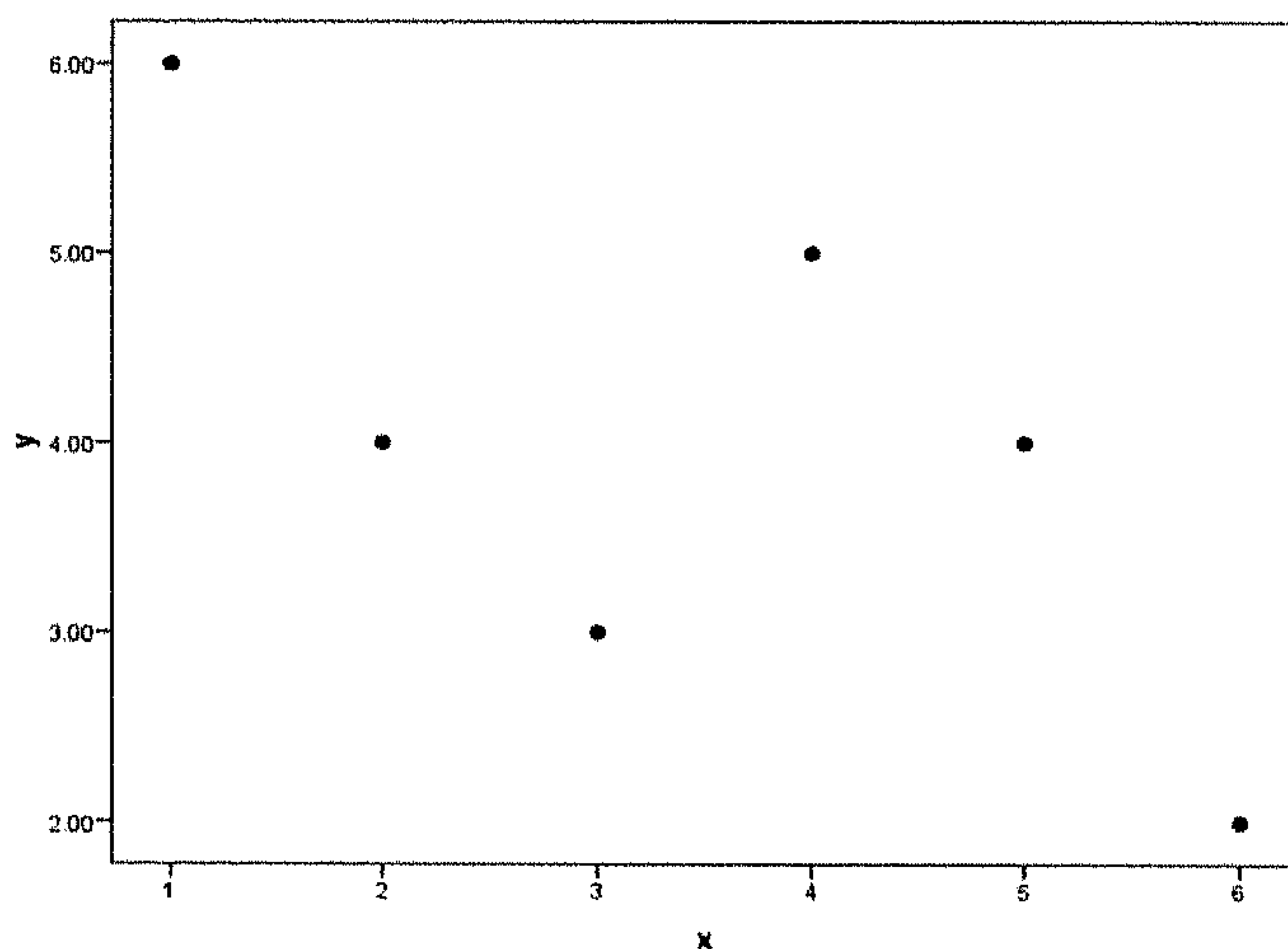


٣- تظهر نافذة جديدة بعنوان Simple Scatter plot



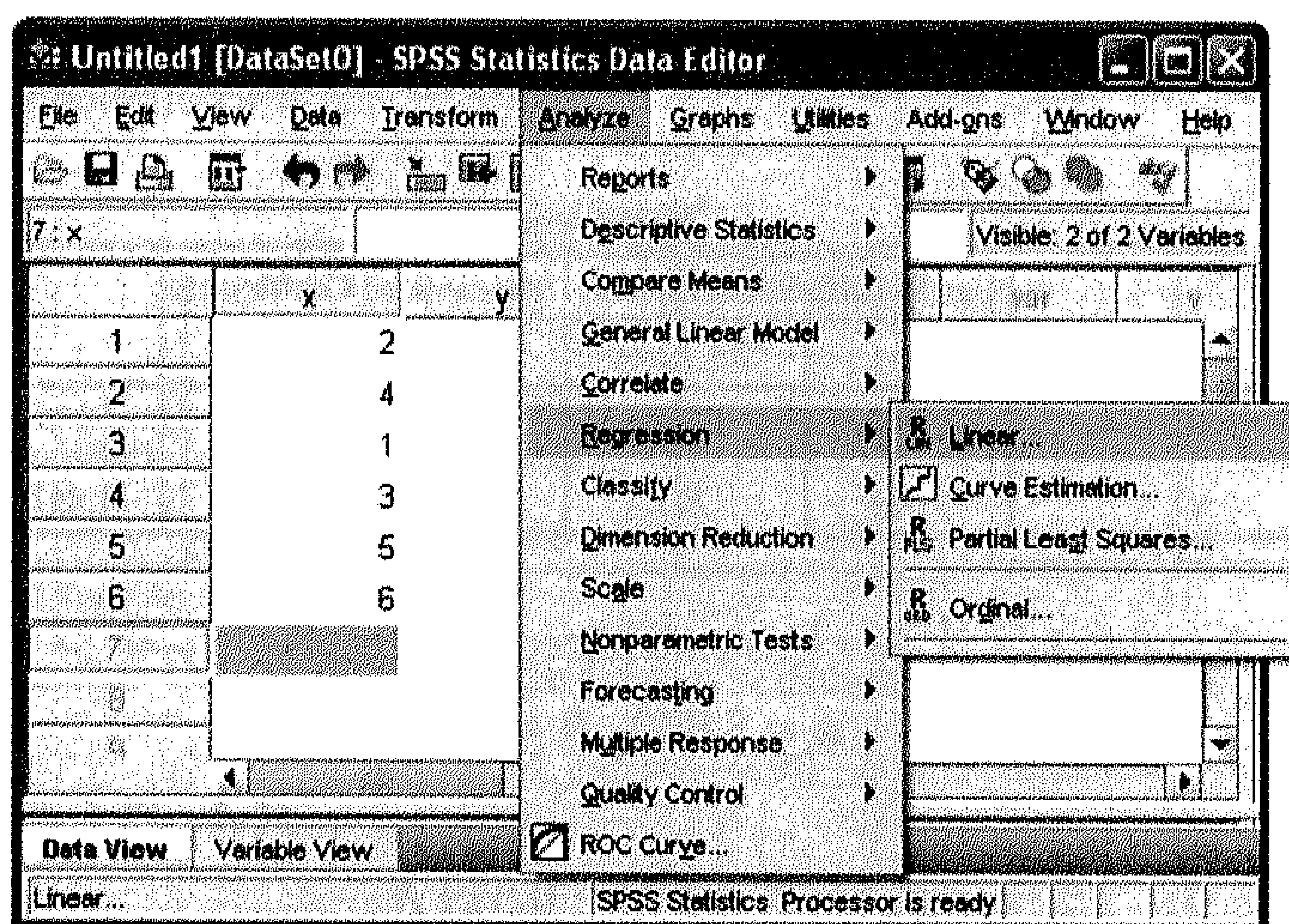
٤- ننقل المتغير X لخانة :X Axis وأيضا ننقل المتغير Y لخانة :Y Axis

٥- ثم نضغط على Ok فننتقل لنافذة المخرجات Output View ويظهر بها الشكل البياني التالي والذي يمثل شكل الانتشار



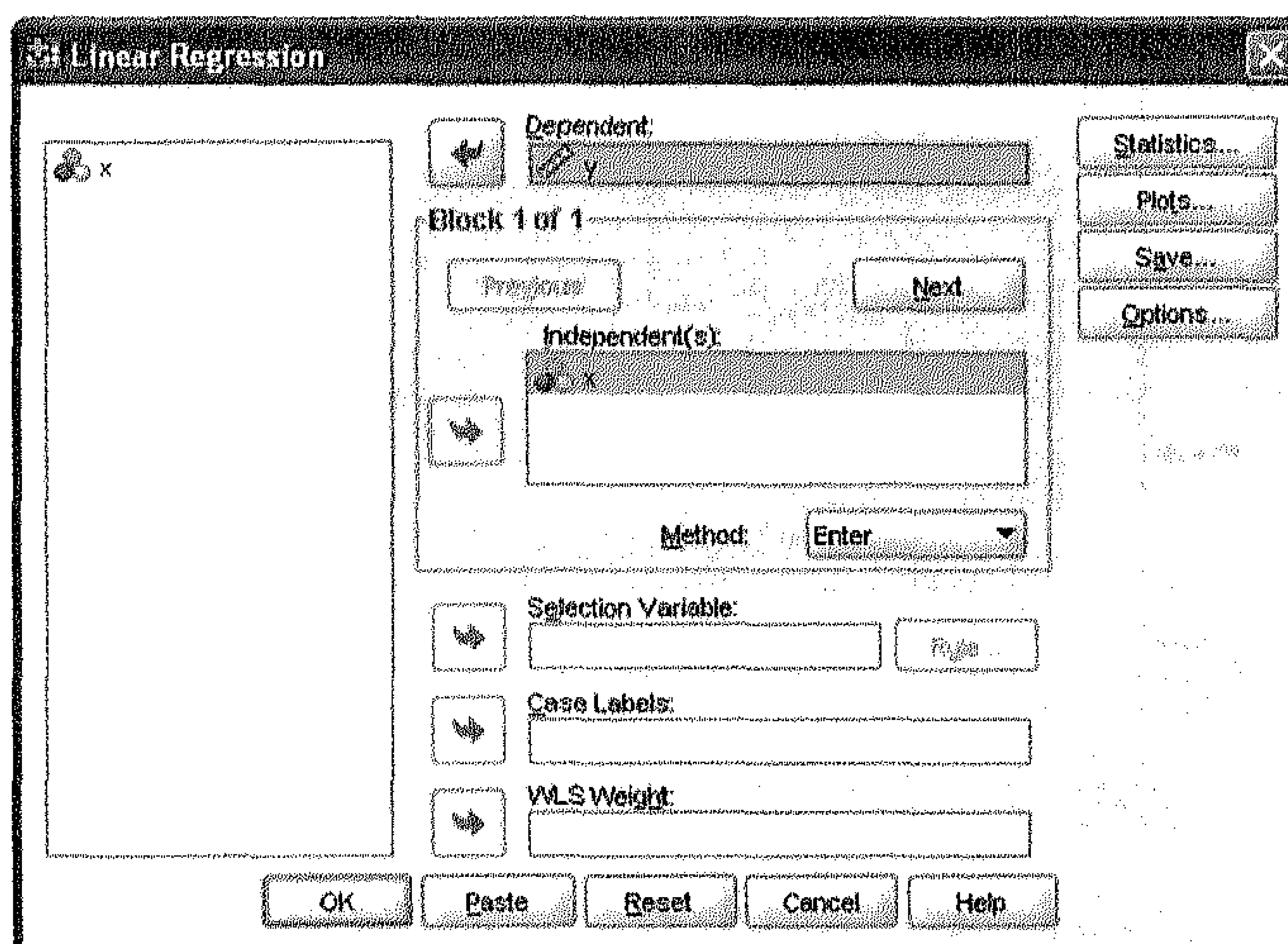
ثالثاً: تعيين خط انحدار Y على X

١- من قائمة Analyze نختار Regression



٢- من القائمة المنسدلة نحدد نوع الانحدار وهو انحدار خطى وسوف نختار Linear

٣- سوف تظهر شاشة جديدة بعنوان Linear Regression



٤- ننقل المتغير X لقائمة Independent(s) والمتغير Y لخانة Dependent

٥- ثم نضغط على Ok فننتقل لنافذة المخرجات والتي تحتوي على عدة جداول ولكننا سوف نهتم هنا بالجدول الذي له العنوان Coefficients التالي:

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.800	1.079		5.376	.006
	x	-.514	.277	-.680	-1.857	.137

a. Dependent Variable: y

من الجدول السابق نجد من العمود Unstandardized Coefficients فنجد أن قيمة الثوابت a, b هي

$$a = 5.800,\ \ b = -0.514$$

وبالتالي فإن خط انحدار Y على X هو

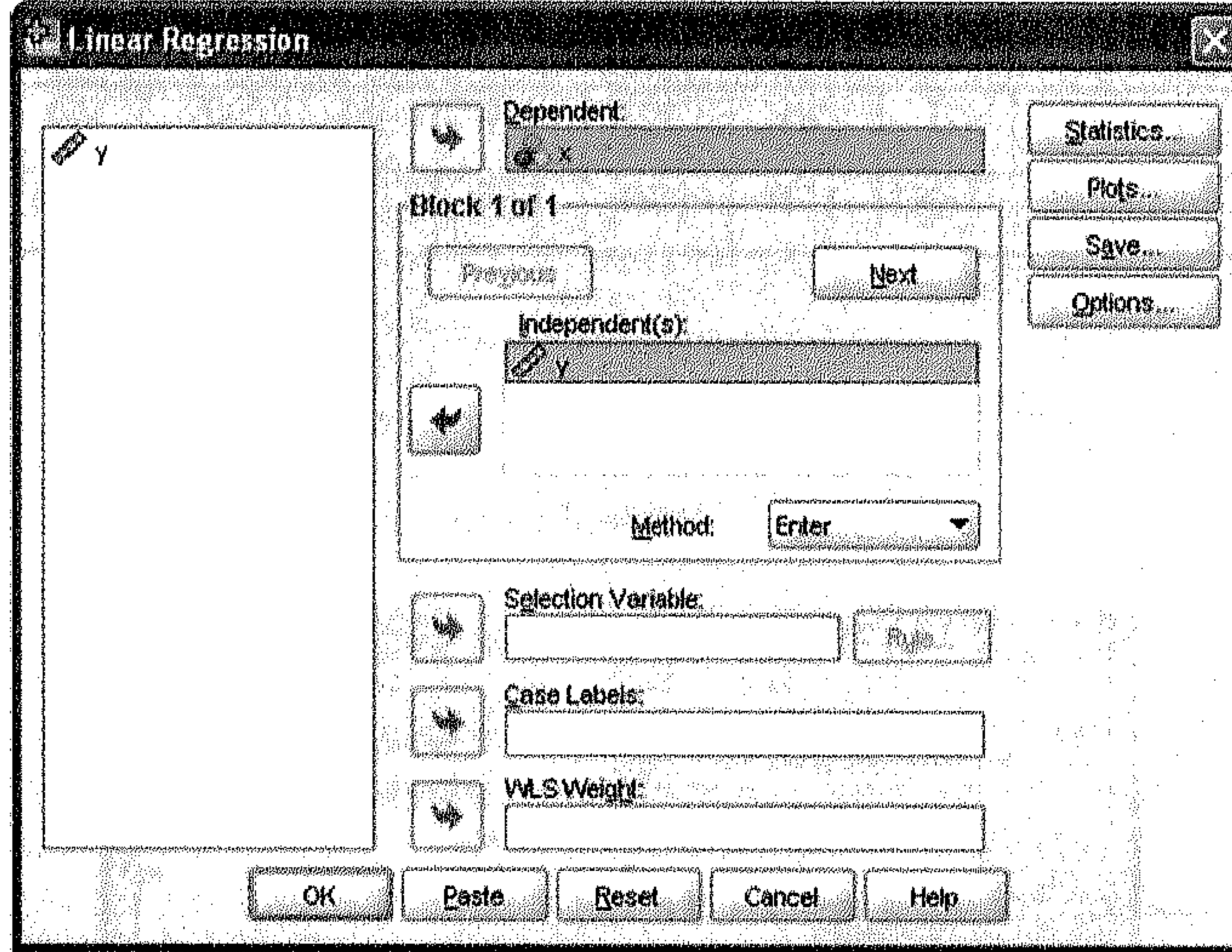
$$Y = 5.8 - 0.514\,X$$

وبالتعويض عن $X = 3.5$ في المعادلة التي حصلنا عليها سوف نحصل على تقدير لقيمة Y المناظرة

$$Y = 5.8 - 0.514(3.5) = 4.001$$

رابعاً: تعيين خط انحدار X على Y

سوف نتبع نفس الخطوات السابقة لكن سيكون التغيير فقط في المتغير المستقل Independent سيكون في هذه الحالة هو Y وسيكون المتغير التابع Dependent هو X



ثم نضغط على Ok فتظهر عدة جداول. ولكننا سوف نهتم هنا بجدول Coefficients التالي

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	7.100	2.038		3.485	.025
	y	-.900	.485	-.680	-1.857	.137

a. Dependent Variable: x

من الجدول السابق نجد من العمود Unstandardized Coefficients نجد أن قيمة الثوابت هي:

$$c = 7.1, \ \ d = -0.90$$

فإن خط انحدار X على Y هو

$$X = 7.1 - 0.90\,Y$$

وبالتعويض عن $Y = 2.3$ في المعادلة السابقة نحصل على تقدير لقيمة X المناظرة

$$X = 7.1 - 0.90(2.3) = 5.03$$

تطبيق(٥-٤)

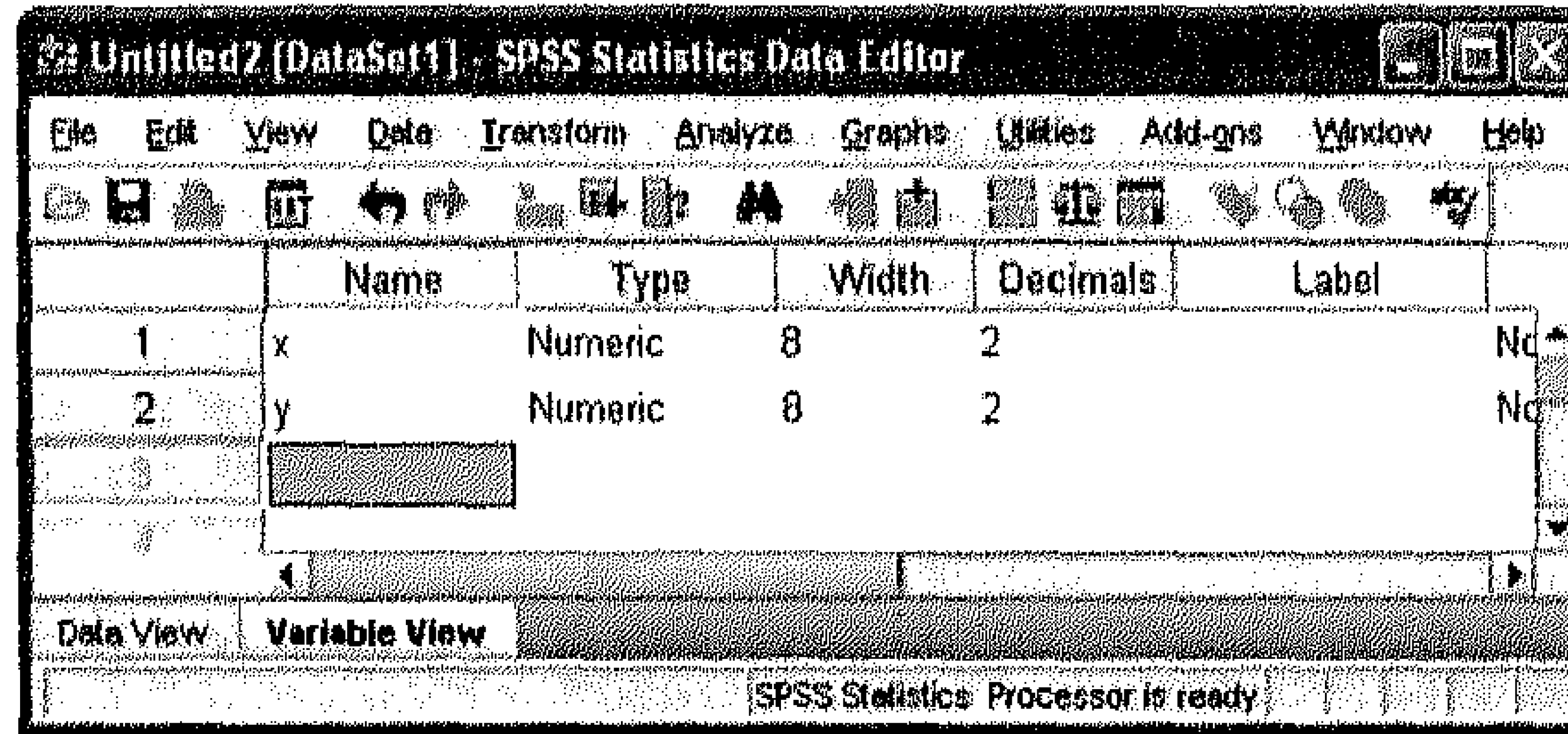
في مثال (٥-٦):احسب معامل الارتباط الخطى لبيرسون بين الطول بالسم (Y) والوزن بالكيلوجرام (X) للبيانات التالية:

الوزن X	61	70	72	65	83	56	62
الطول Y	165	170	170	169	170	154	164

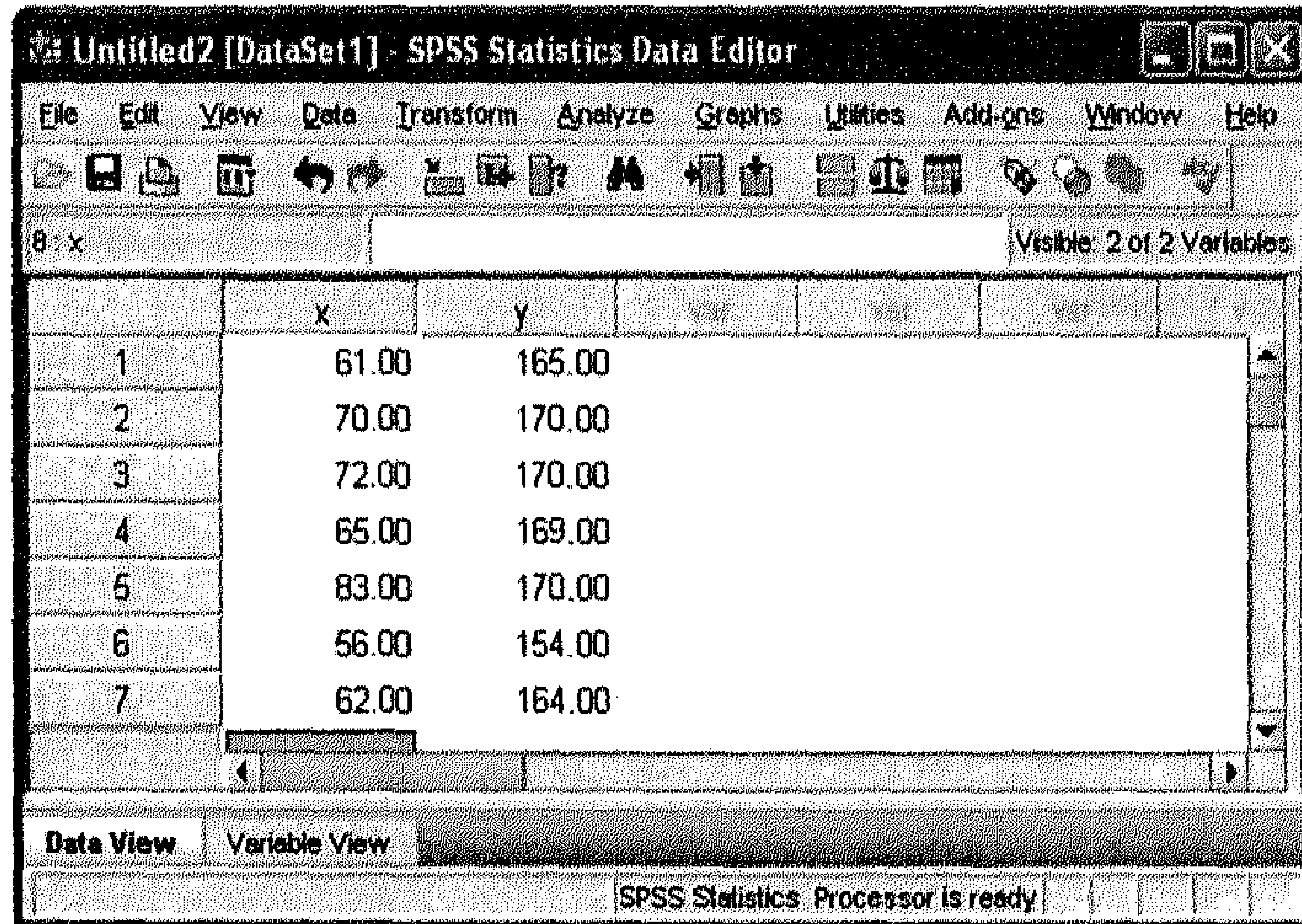
الحل:

أولاً: إدخال البيانات

١- نقوم بتعريف المتغيرين X, Y في نافذة Variable View

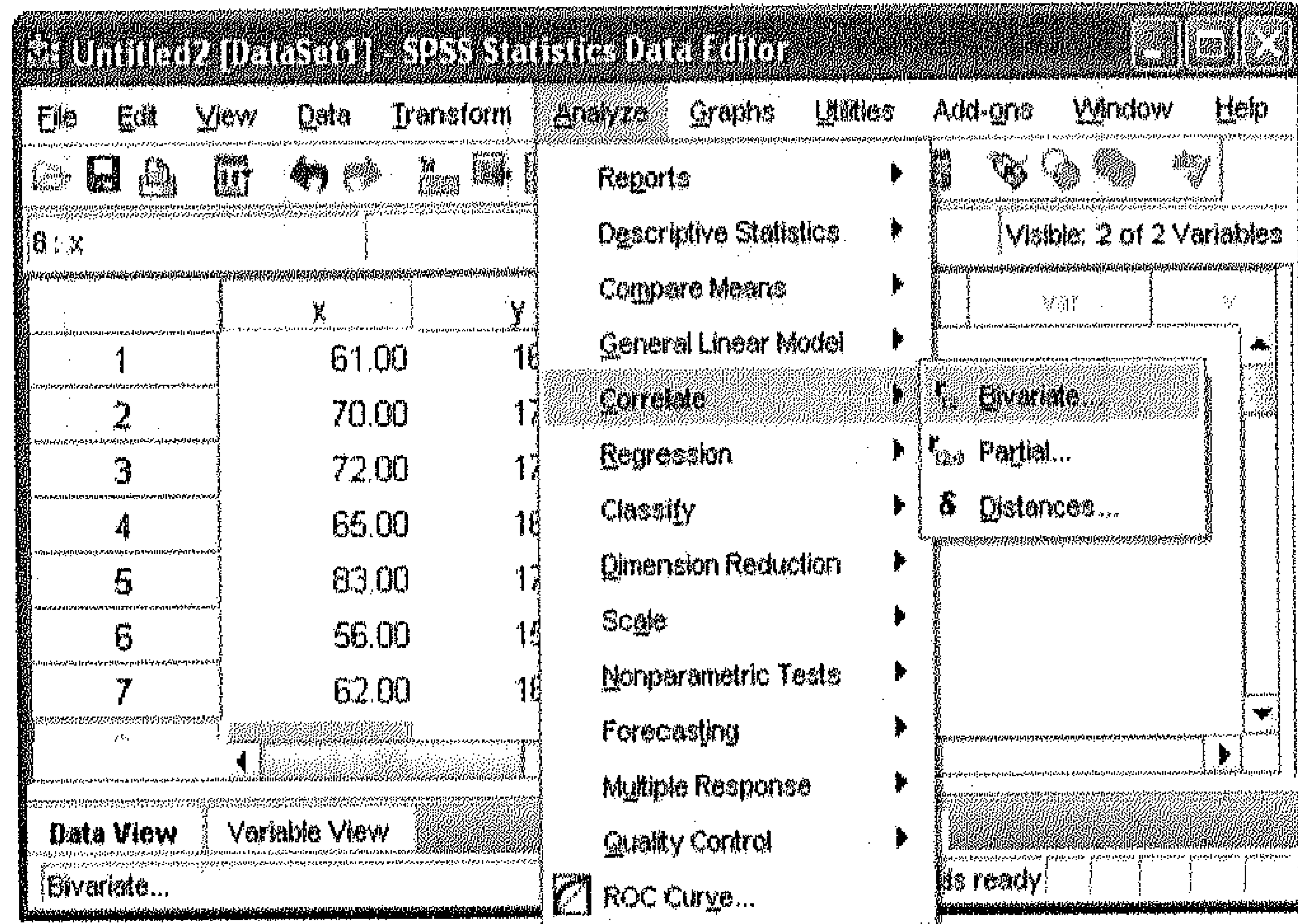


١- ثم نتنقل لنافذة Data View ونقوم بإدخال البيانات

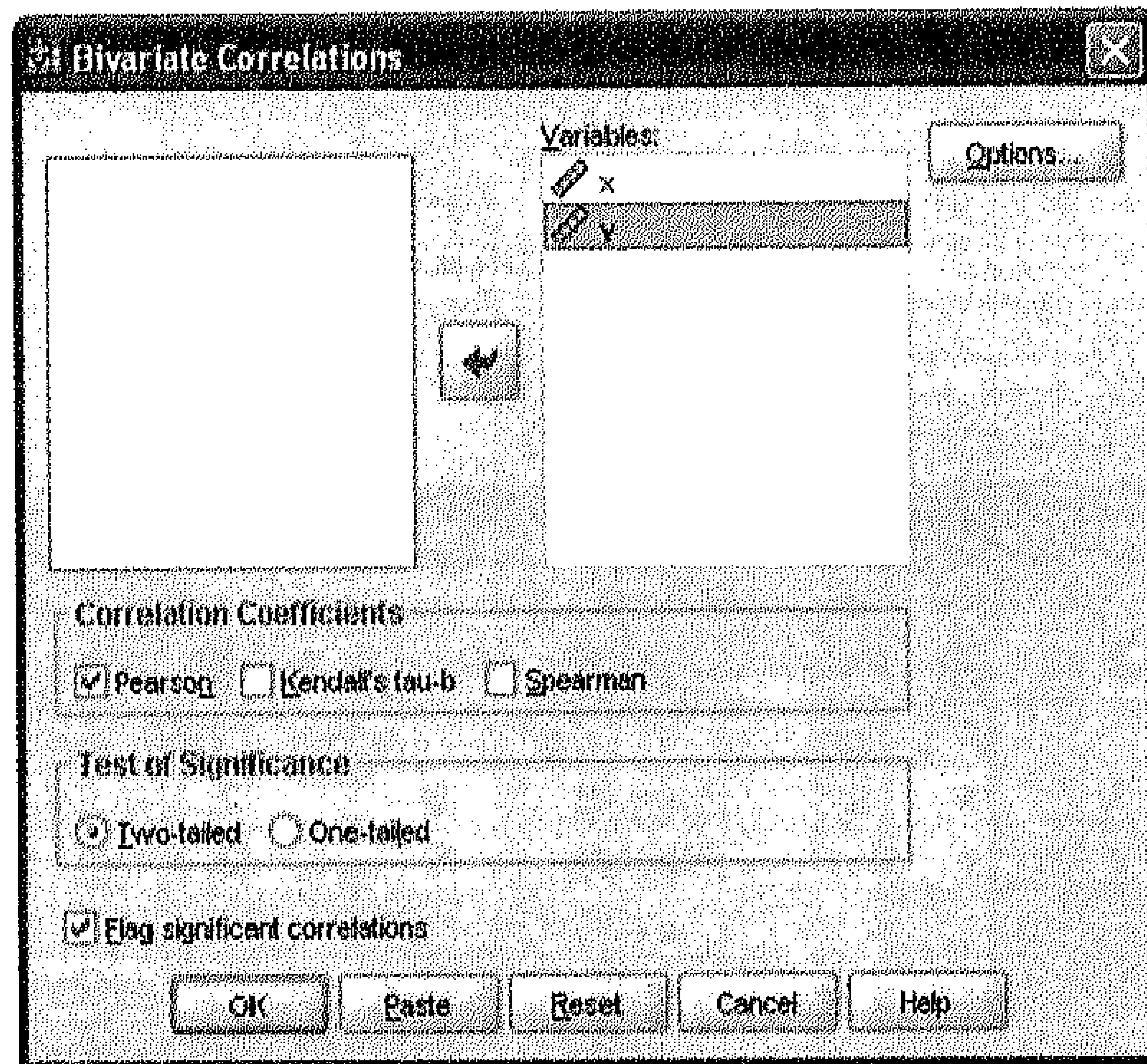


ثانياً: تعيين معامل الارتباط

١- نضغط علي قائمة Analyze ونختار Correlate فتظهر قائمة منسدلة فنختار منها Bivariate



٢- تظهر شاشة جديدة بعنوان Bivariate Correlations



٣- فننقل المتغيرين X, Y لقائمة Variable(s) ونحدد نوع معامل الارتباط Pearson ثم نضغط Ok فننتقل لنافذة المخرجات Output view ونحصل على النتائج التالية:

Correlations

		x	y
x	Pearson Correlation	1	.760*
	Sig. (2-tailed)		.047
	N	7	7
y	Pearson Correlation	.760*	1
	Sig. (2-tailed)	.047	
	N	7	7

*. Correlation is significant at the 0.05 level (2-tailed).

ونجد من الجدول السابق أن معامل ارتباط بيرسون بين X,Y يساوى 0.76 وهو ارتباط طردي قوى.

تطبيق(٥-٥)

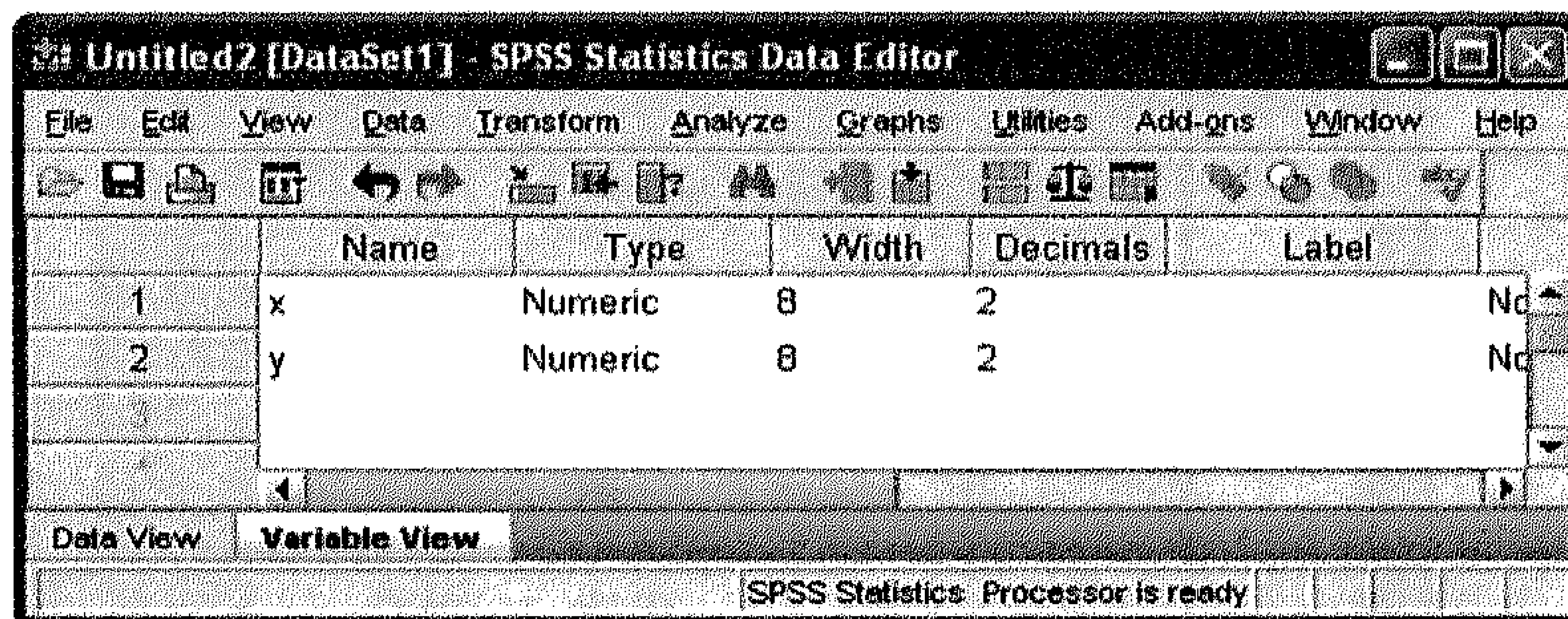
في مثال (٥-٩): فيما يلي درجات 12 طالب في اختبار مقرري 1080 ريض (X)، 1080 حيا (Y) والمطلوب تعيين معامل ارتباط سبيرمان للرتب بين تقديرات المقررين.

X	16	12	10	13	9	10	11	13	14	17	19	20
Y	20	11	14	16	15	10	17	11	12	20	18	12

الحل

أولاً: نقوم بإدخال البيانات للبرنامج وذلك باتباع الخطوات التالية:

١- نقوم بتعريف المتغيرات X,Y من نافذة Variable View وهى متغيرات كمية Numeric



٢- ثم ننتقل لنافذة Data View ونقوم بإدخال قيم المتغيرات

Untitled1 [DataSet0] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

13 : y — Visible: 2 of 2 Variables

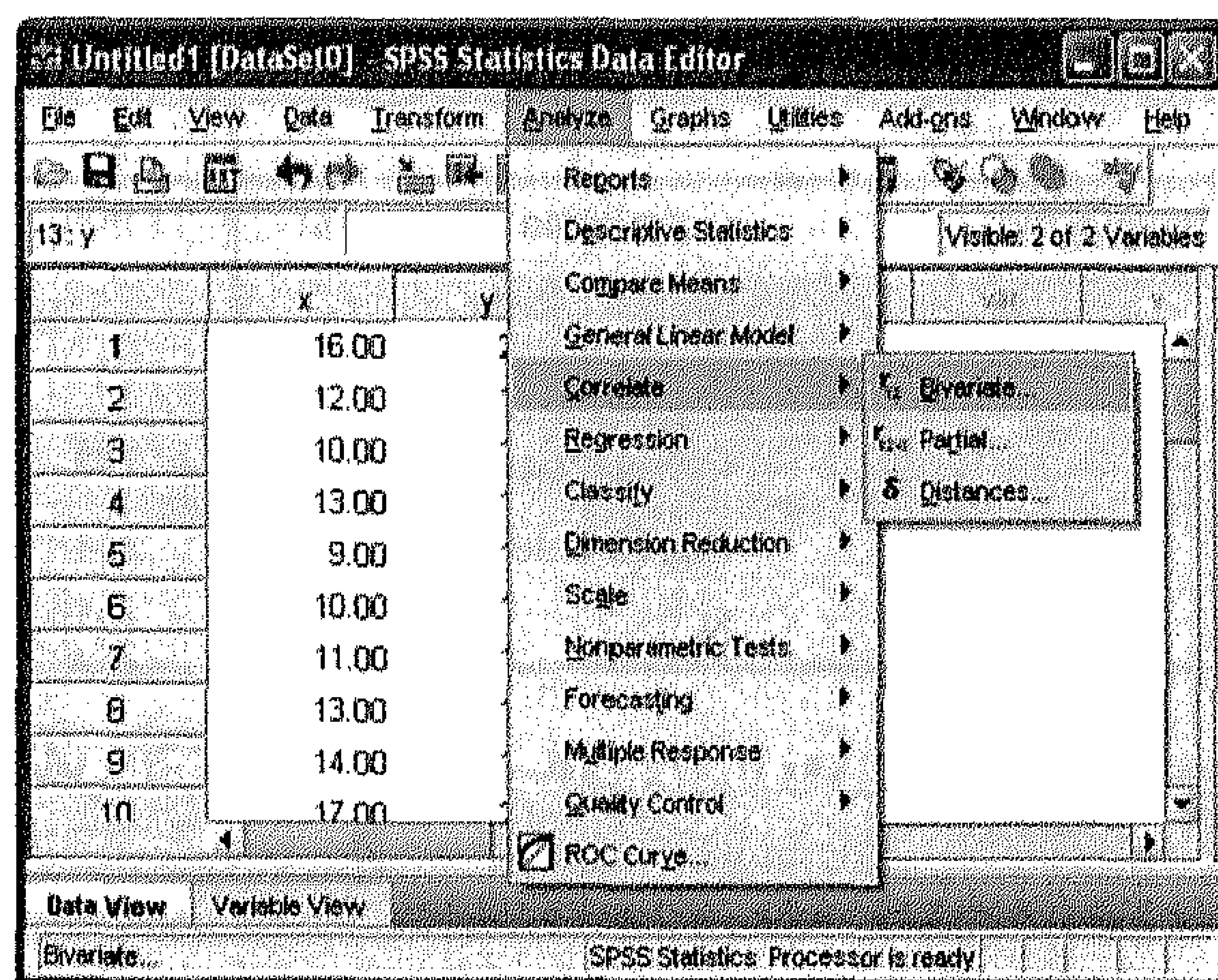
	x	y	var	var	var
1	16.00	20.00			
2	12.00	11.00			
3	10.00	14.00			
4	13.00	16.00			
5	9.00	15.00			
6	10.00	10.00			
7	11.00	17.00			
8	13.00	11.00			
9	14.00	12.00			

Data View | Variable View

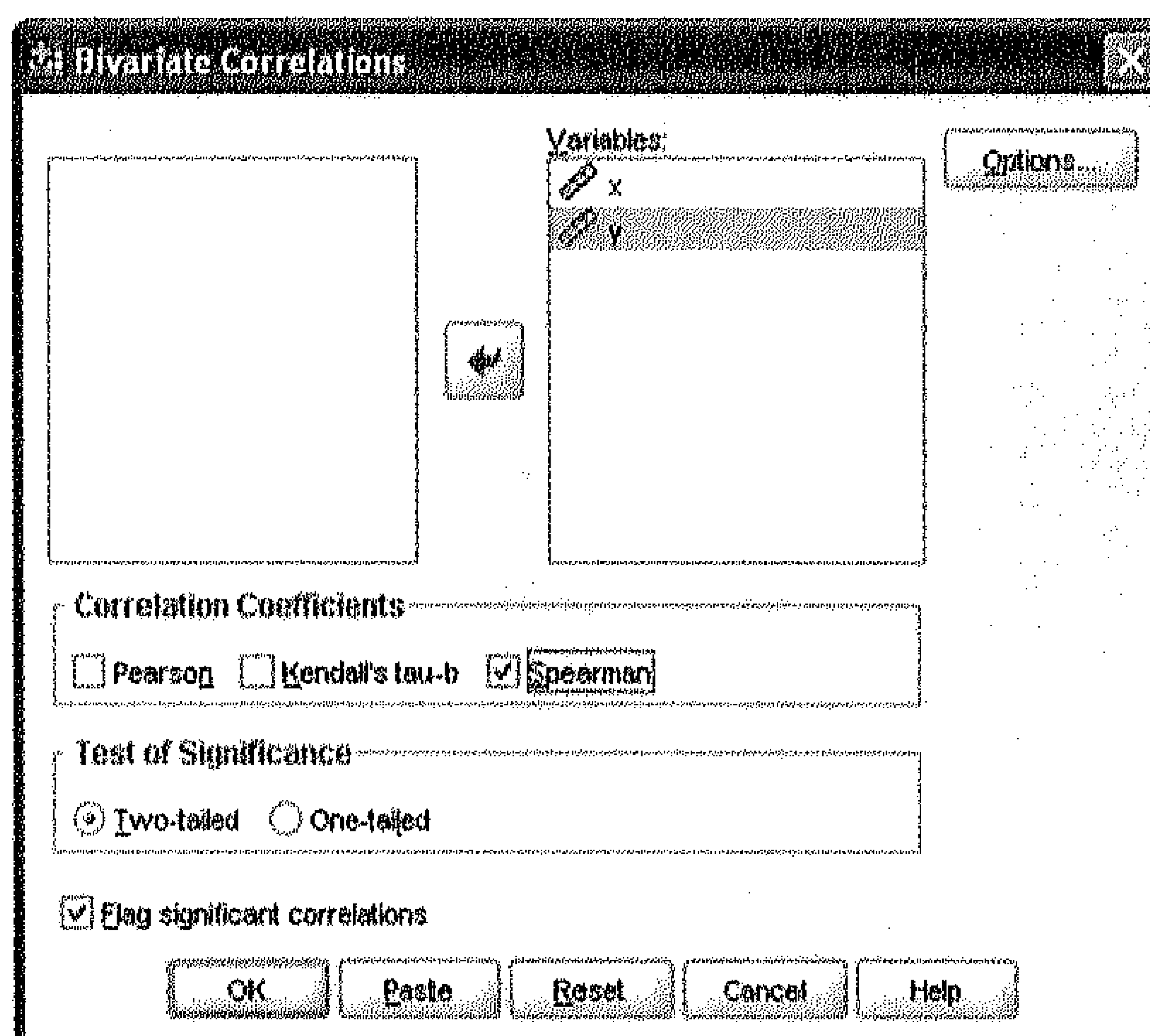
SPSS Statistics Processor is ready

ثانياً: لتعيين معامل الارتباط نتبع الخطوات التالية:

١- من قائمة Analyze نختار Correlate



٢- من القائمة المنسدلة نختار Bivariate فتظهر نافذة جديدة بعنوان Bivariate Correlations



٣- ننقل المتغيرين X, Y لقائمة Variable(s) ونحدد نوع معامل الارتباط Spearman ثم نضغط Ok فننقل لنافذة المخرجات Output view وتظهر النتائج التالية

Correlations

			x	y
Spearman's rho	x	Correlation Coefficient	1.000	.372
		Sig. (2-tailed)	.	.234
		N	12	12
	y	Correlation Coefficient	.372	1.000
		Sig. (2-tailed)	.234	.
		N	12	12

ونجد من الجدول السابق أن معامل الارتباط بين X,Y يساوي 0.372 وهو ارتباط طردي ضعيف.

تطبيق(٥-٦)

في مثال (٥-١١): احسب معامل التوافق بين لون الشعر ولون العيون لعينة من 45 فرداً والتي لخصت بياناتهم كما في الجدول المزدوج التالي:

لون الشعر / لون العين	أشقر	بني	أسود	المجموع
أزرق	6	5	4	15
عسلي	3	6	6	15
أسود	2	7	6	15
المجموع	11	18	16	45

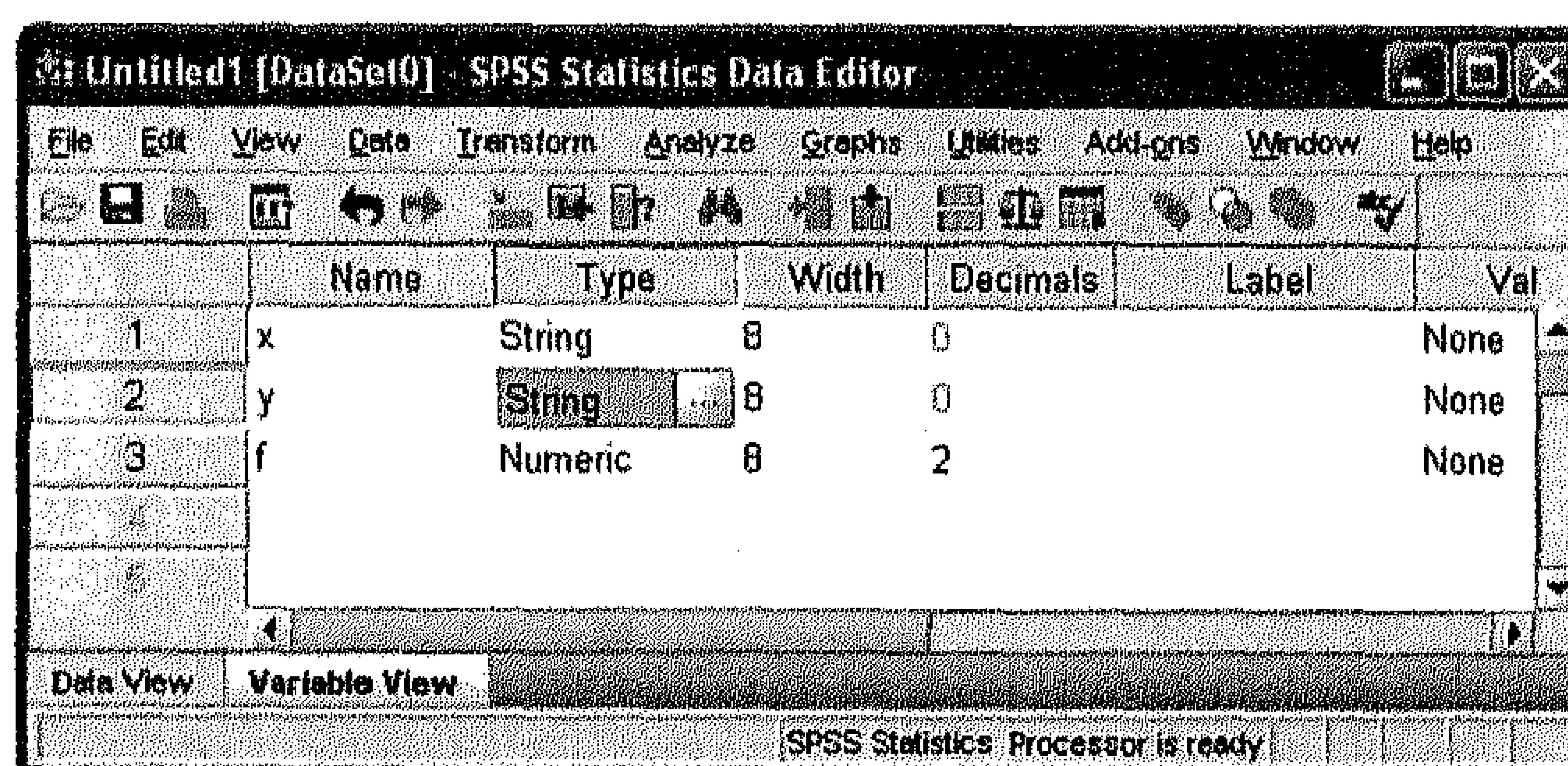
الحل:

الإصدار 17 من برنامج SPSS لا يقبل التعامل مع البيانات المكتوبة باللغة العربية لذا سوف نترجم الكلمات العربية للغة الانجليزية ونتعامل معها. بحيث

- متغير لون الشعر سوف نستخدم: أسود تعني Black، أشقر تعني yellow، بني تعني brown
- متغير لون العين سوف نستخدم: أزرق تعني blue، أسود تعني black، عسلي تعني Bbrown

أولاً: لإدخال البيانات للبرنامج سوف نتبع الخطوات التالية

١- من نافذة Variable View نقوم بتعريف ثلاثة متغيرات X, Y, f وهى تمثل قيم المتغير X لون الشعر، Y لون العين، f التكرار المشاهد



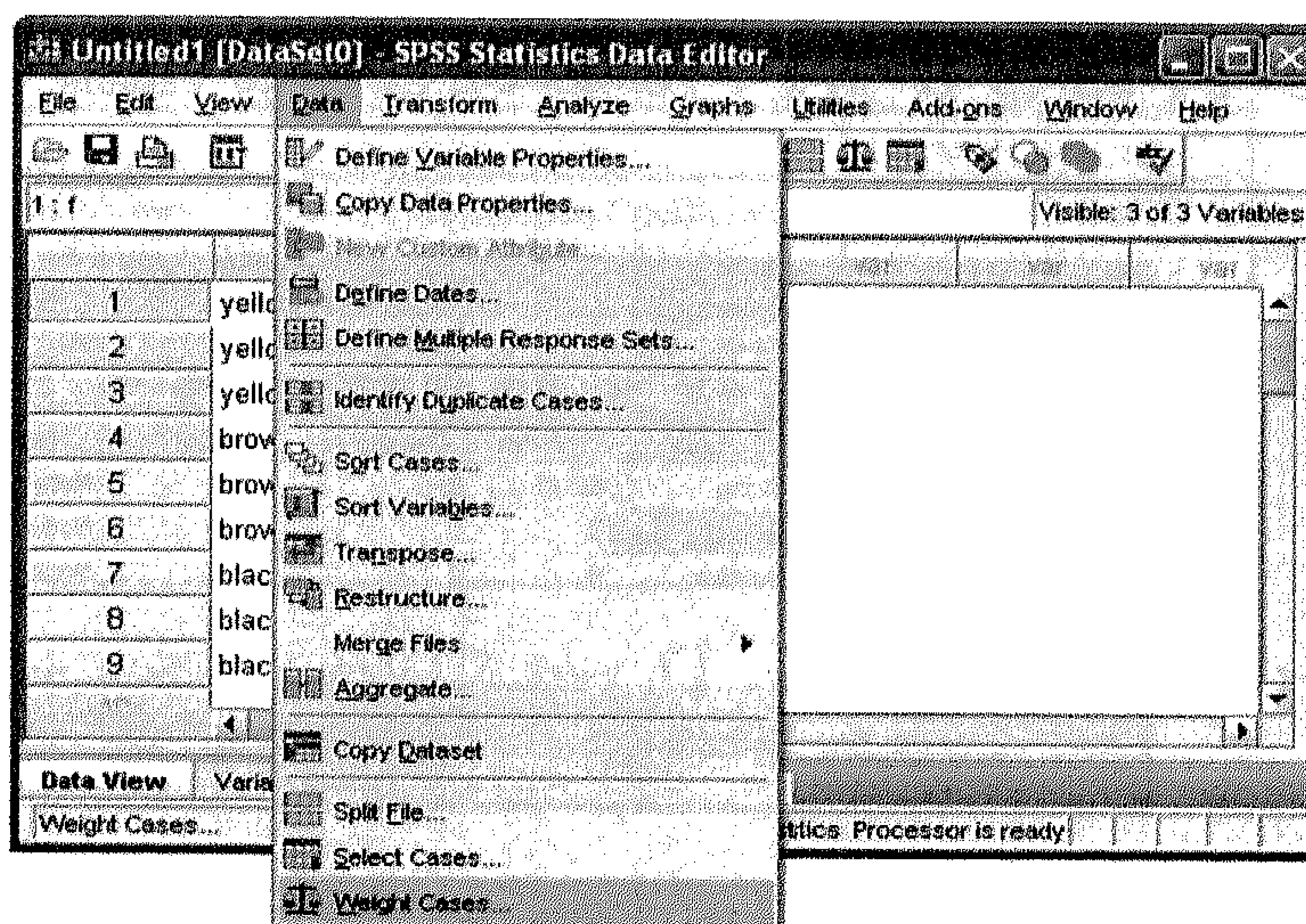
٢- من نافذة Data View نقوم بإدخال البيانات كما يلي:

Untitled1 [DataSet0] - SPSS Statistics Data Editor

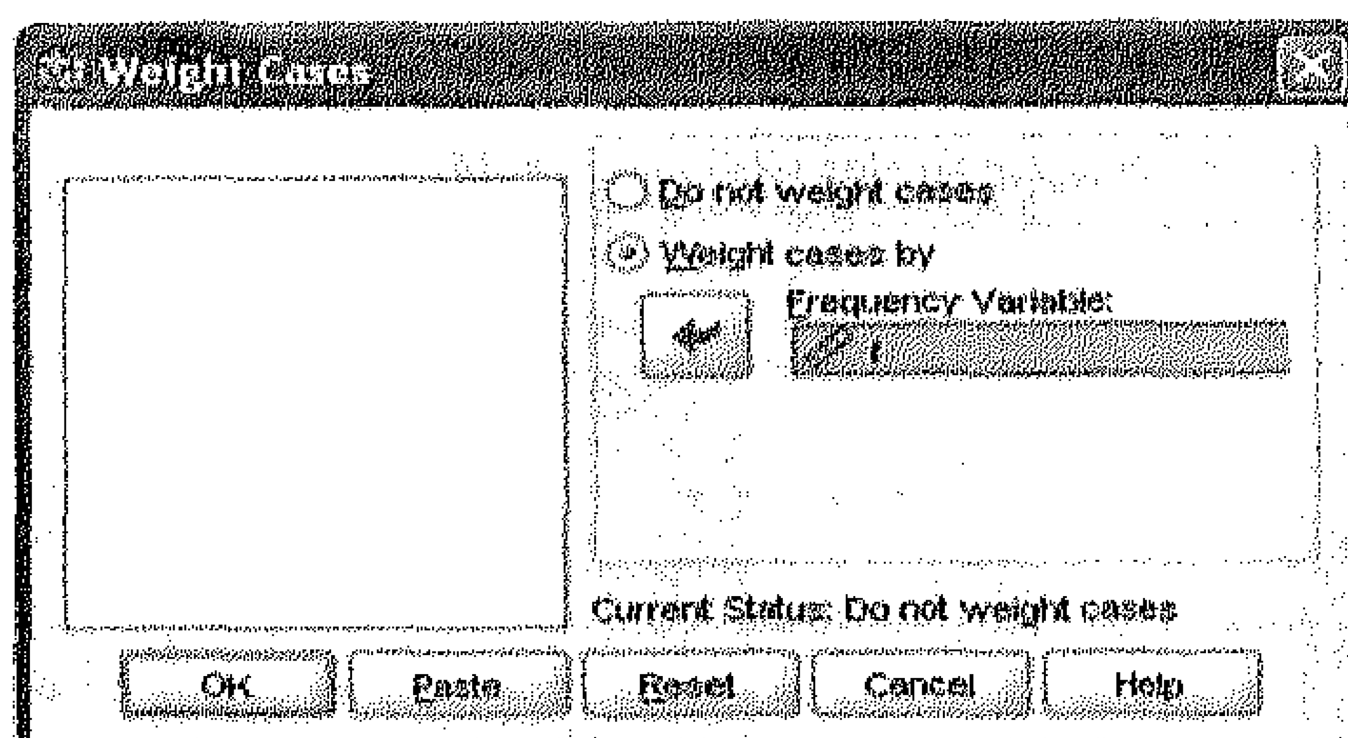
	x	y	f
1	yellow	blue	6.00
2	yellow	bbrown	3.00
3	yellow	black	2.00
4	brown	blue	5.00
5	brown	bbrown	6.00
6	brown	black	7.00
7	black	blue	4.00
8	black	bbrown	6.00
9	black	black	6.00

ثانياً: تعريف أن f هي تكرار للمتغيرين X, Y عن طريق

١- من قائمة Data نختار Weight Cases



٢- تظهر شاشة جديدة بعنوان weight Cases نختار Weight case by

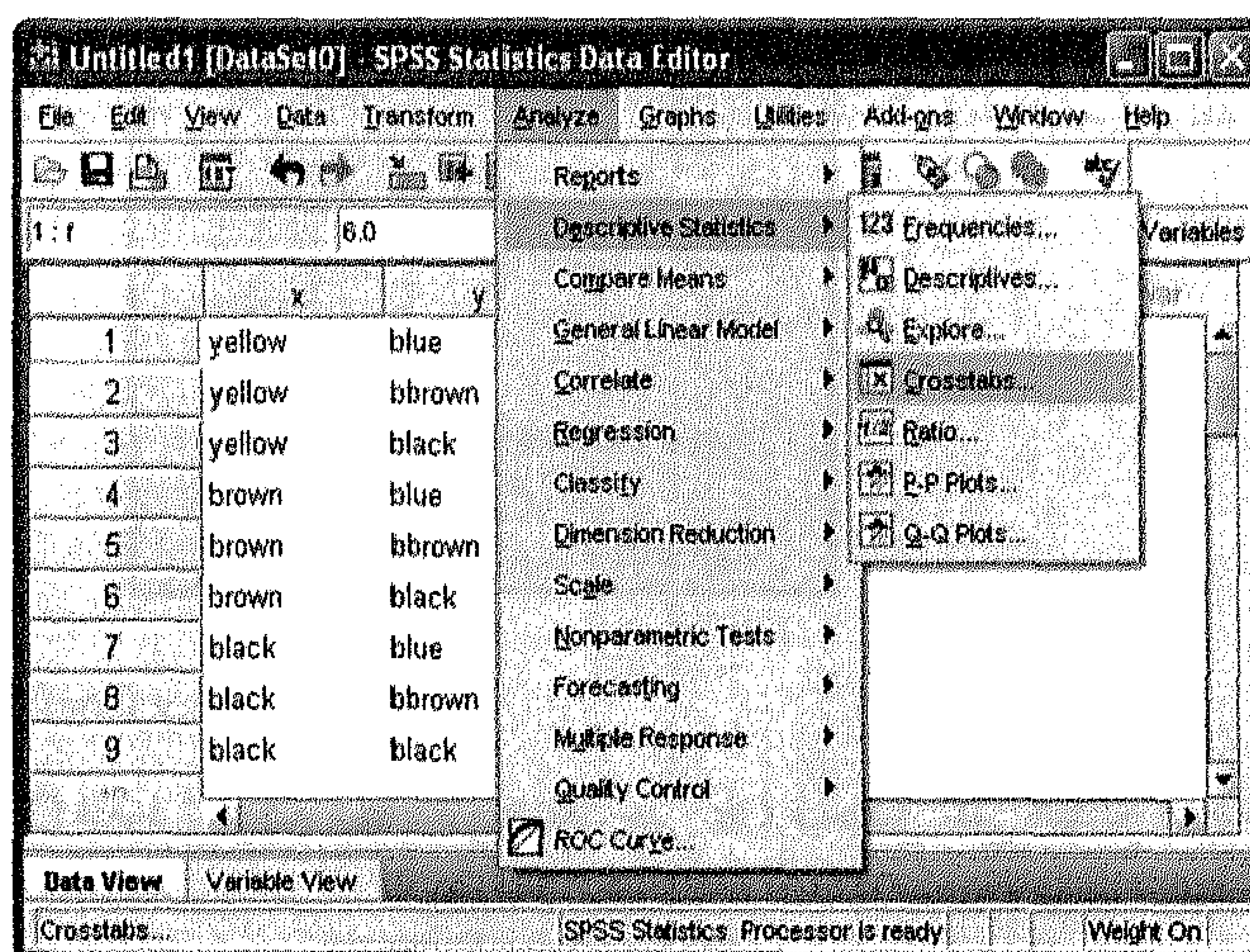


٣- ثم ننقل المتغير f لخانة: Frequency Variable ونضغط Ok فنعود لملف البيانات دون حدوث تغيير ظاهري للبيانات

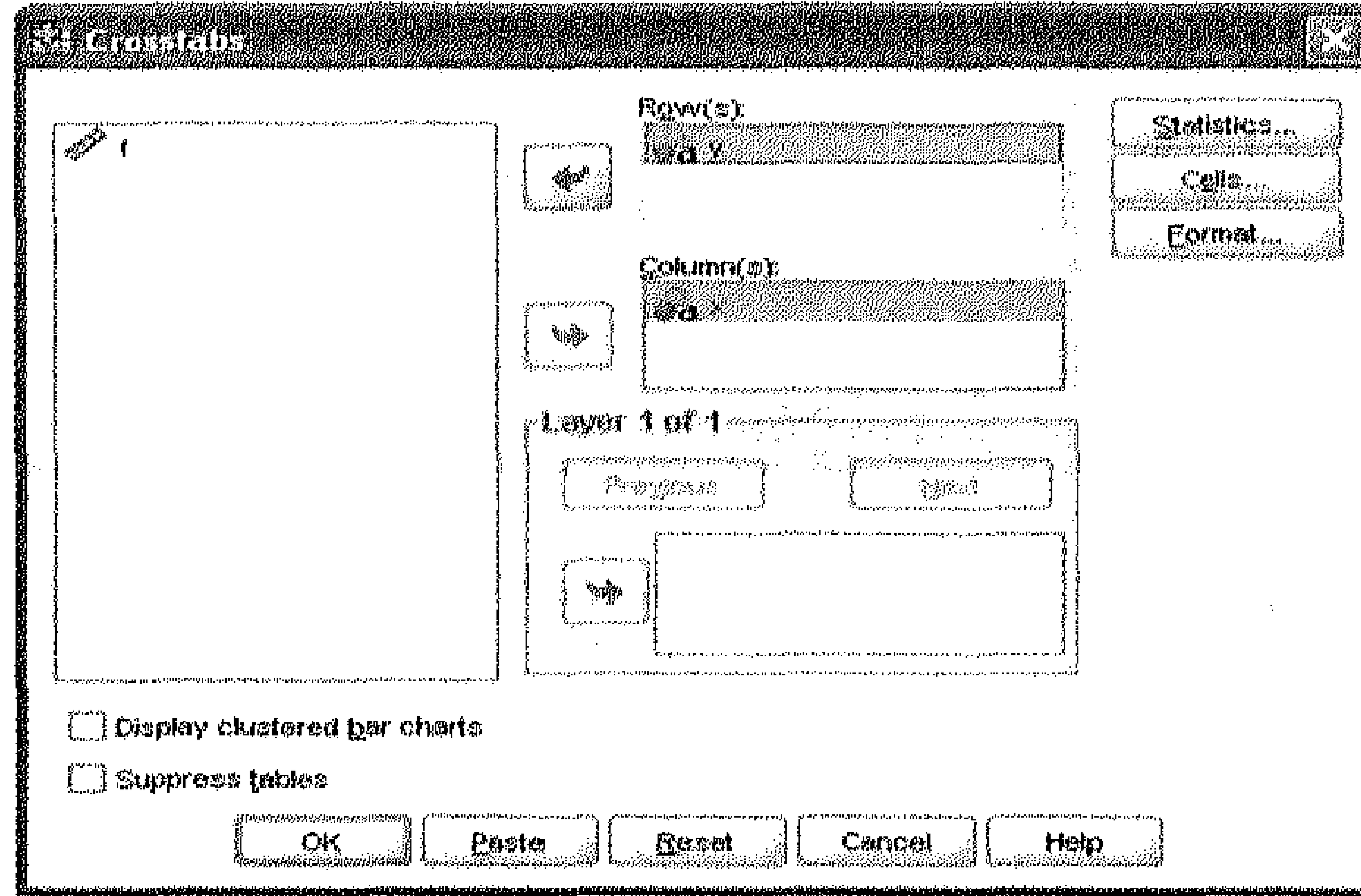
ثالثاً لتعيين معامل التوافق بين لون الشعر ولون العيون نتبع الخطوات التالية:

١- من قائمة Analyze نختار Descriptives Statistics

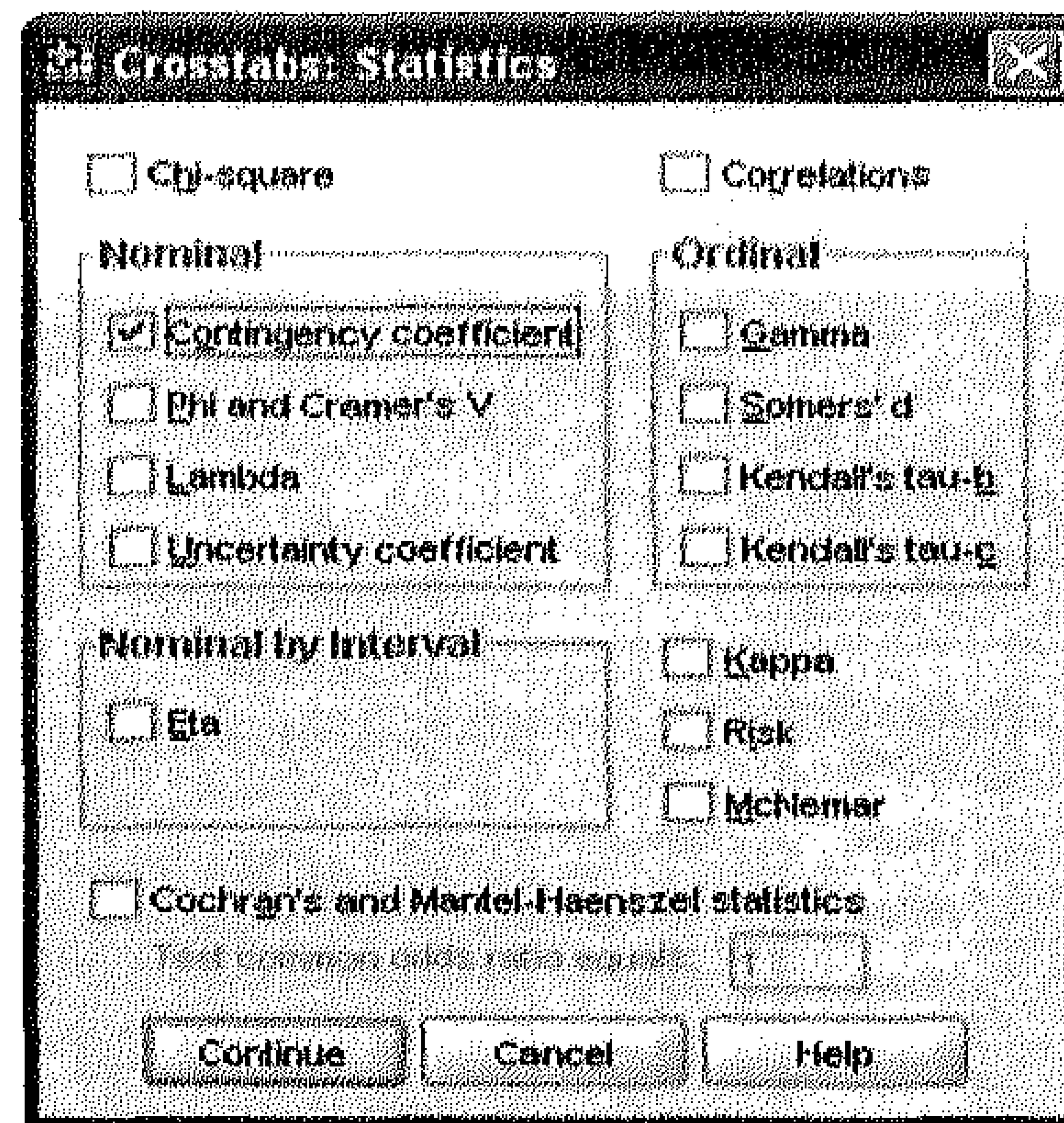
٢- فتظهر قائمة منسدلة نختار منها Crosstabs



٣- فتظهر شاشة جديدة بعنوان Crosstabs ننقل المتغير Y لقائمة Row(s) والمتغير X لقائمة Column



٤- نضغط على Statistics تظهر شاشة جديدة بعنوان Crosstabs: Statistics



٥- نحدد الاختيار Contingency Coefficient

٦- نضغط على Continue فنعود للشاشة السابقة ونضغط على Ok فنحصل على النتائج التالية:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
y * x	45	100.0%	0	.0%	45	100.0%

الجدول الأول ويعطى العدد الكلى لقيم المتغيرين وهو عبارة عن مجموع التكرار وهم 45 قيمة

y * x Crosstabulation

		x			Total
		black	brown	yellow	
y	bbrown	6	6	3	15
	black	6	7	2	15
	blue	4	5	6	15
Total		16	18	11	45

الجدول الثاني ويمثل الجدول التكراري المزدوج للمتغيرين X, Y ويحتوى على التوزيع التكراري الهامشي.

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Contingency Coefficient	.258	.525
N of Valid Cases		45	

الجدول الثالث يعطى قيمة معامل التوافق (معامل الاقتران في حالة أكثر من صفتين للظاهرة) ونجد أن قيمة هي 0.258 وهى ضعيفة.

أسئلة وتمارين (٥)

١- أي القيم التالية لا يمثل قيمة لمعامل الارتباط

(i) -0.31 (ii) 0.977 (iii) 1.01 (iv) -0.92

٢- أي من القيم التالية يمكن أن يمثل معامل ارتباط بين متغيرين

(i) 2 (ii) $\frac{4}{3}$ (iii) 0 (iv) لا شئ مما سبق

٣- أي الأعداد التالية يمثل معامل ارتباط عكسياً

(i) - 1.31 (ii) 0.977 (iii) 1.01 (iv) -0.92

٤- أي من القيم التالية يمثل معامل ارتباط طردياً قوياً

(i) -0.31 (ii) 0.977 (iii) 1.01 (iv) 0.592

٥- إذا كان عدد الأزواج للظاهرتين X, Y يساوى 10 وكانت جميع قيم الظاهرة X متساوية، وجميع قيم الظاهرة Y متساوية فإن معامل الارتباط لسبيرمان بين الظاهرتين (X,Y) يساوى

(i) 0 (ii) -1 (iii) 1 (iv) 0.5

٦- إذا كان جميع قيم المتغير X متساوية وجميع قيم Y أيضا متساوية، وكان لدينا 20 زوجاً من القيم للمتغيرين (X,Y) فإن معامل ارتباط سبيرمان للرتب هو

(i) 0 (ii) 1 (iii) 0.999 (iv) غير ذلك

٧- إذا كانت معادلة انحدار Y على X هي $Y = 0.6X + 12$ فإن القيمة التقديرية لـ Y عندما $X = 20$ هي

(i) 12 (ii) 24 (iii) 32 (iv) غير ذلك

٨- أوجد خط انحدار Y على X لمجموعة البيانات التالية:

X	0	1	2	3	4	5	6	7	8
Y	160	168	176	179	183	186	189	186	184

ثم قدر قيمة Y عندما $X = 4.5$

٩- البيانات التالية تمثل استهلاك الوقود (بمليون جالون) في أحد البلاد في الفترة من عام 1977 حتى عام 1983

العام X	1977	1978	1979	1980	1981	1982	1983
الاستهلاك Y	32.5	37.1	35.5	37.7	41.5	46.4	44.8

(أ) ارسم شكل الانتشار

(ب) أوجد أفضل خط مستقيم على الصورة $Y = a + bX$ والذى يقرب البيانات. ومنه قدر قيمة الاستهلاك عام 1984

١٠- أراد احد الباحثين دراسة العلاقة بين عدد أفراد الأسرة (X) وعدد السيارات التي تمتلكها الآسرة (Y) فقام باختيار 25 أسرة، وتم تجميع البيانات كالتالي:

عدد إفراد الأسرة X	4	6	4	3	2	5	6	4	5	6
عدد السياراتY	2	3	3	2	1	3	2	1	3	3
عدد إفراد الأسرة X	3	4	5	4	3	2	2	4	5	4
عدد السياراتY	2	1	3	2	2	1	1	3	3	1
عدد إفراد الأسرة X	5	6	5	4	6	3	4	3	2	5
عدد السياراتY	1	2	3	4	5	1	3	1	1	2

كون الجدول التكراري المزدوج لعدد إفراد الأسرة (X) وعدد السيارات التي تمتلكها الأسرة (Y)

١١- ارسم شكل الانتشار ثم أوجد خط انحدار X على Y لمجموعة البيانات التالية:

X	25	30	35	40	45	50
Y	78	70	65	58	48	42

ثم قدر قيمة X عندما $Y = 54$

١٢- الجدول التالي يحتوى على درجات 10 طلاب في الاختبار الأول (X) والاختبار الثاني (Y) في مقرر 1040إحص

X	12	14	7	8	16	17	18	20	19	11
Y	10	16	17	10	17	18	20	19	17	16

(أ) ارسم شكل الانتشار.

(ب) إذا حصل الطالب على 15 في الاختبار الأول فما هي درجته تقريباً في الاختبار الثاني.

١٣- الجدول التالي يبين تقديرات 8 طلاب في مقرري الإحصاء والكيمياء

الإحصاء	ضعيف جدا	ممتاز	مقبول	ممتاز	جيد	جيد جدا	ضعيف	جيد
الكيمياء	جيد	جيد جدا	جيد	ممتاز	ضعيف	جيد	جيد	ممتاز

والمطلوب

(أ) حساب معامل الارتباط بين تقديرات الطلاب في المقررين.

(ب) تحديد نوع العلاقة بين تقديرات الطلاب في المقررين.

١٤- احسب معامل الارتباط الخطى لبيرسون بين الطول (X) بالسم والوزن (Y) بالكيلوجرام لمجموعة من طلاب كلية العلوم والدراسات الإنسانية بالخرج

الطول X	160	165	170	165	168	154	160	169	173
الوزن Y	60	73	65	78	80	62	64	63	70

١٥- احسب معامل الاقتران بين التعليم والتدخين لمجموعة من الأفراد من سكان مدينة الرياض حيث تم تلخيص البيانات التي تم تجميعها في الجدول المزدوج التالي:

التعليم / التدخين	غير متعلم	متعلم
مدخن	9	15
غير مدخن	11	5

١٦- إذا علمت أن الوسط الحسابي والانحراف المعياري لدرجات الطلاب في الاختبار الأول لمقرر 1020 ريض (X) هما $\bar{X} = 14, S_X = 2.4$ والمتوسط والانحراف المعياري لدرجات الطلاب في الاختبار الثاني (Y) هما $\bar{Y} = 15, S_Y = 2$

(أ) فأوجد معادلة انحدار Y على X علماً بأن الارتباط بين X, Y هو 0.459

(ب) قدر درجة الطالب في الاختبار الثاني إذاكانت درجته في الاختبار الأول هي 14

١٧- ارسم شكل الانتشار لمجموعة البيانات التالية وحدد نوع العلاقة بينهما

X	2	4	7	3	5	6	8	10
Y	5	13	5	11	17	20	24	31

ثم أوجد

(أ) معامل الارتباط الخطى.

(ب) معادلة خط انحدار X على Y

١٨- عند دراسة العلاقة بين الدخل والجنسية لعينة من 40 فرداً من سكان محافظة الخرج بالمملكة العربية السعودية حصلنا على النتائج التالية:

الجنسية / الدخل	سعودي	غير سعودي	المجموع
منخفض	4	7	11
متوسط	6	5	11
مرتفع	12	6	18
المجموع	22	18	40

احسب معامل التوافق بين مستوى الدخل والجنسية.

١٩- كون الجدول التكراري المزدوج لمتغيري الجنسية (X) ومستوى الدخل(Y)

الدخلY	الجنسية X	الدخلY	الجنسية X	الدخلY	الجنسية X
مرتفع	غير سعودي	متوسط	سعودي	ضعيف	سعودي
متوسط	سعودي	ضعيف	سعودي	متوسط	سعودي
ضعيف	سعودي	متوسط	سعودي	مرتفع	غير سعودي
متوسط	غير سعودي	مرتفع	غير سعودي	مرتفع	غير سعودي
متوسط	سعودي	مرتفع	غير سعودي	متوسط	غير سعودي
مرتفع	سعودي	متوسط	غير سعودي	ضعيف	سعودي
ضعيف	غير سعودي	ضعيف	سعودي	ضعيف	غير سعودي
		ضعيف	غير سعودي	متوسط	سعودي
		متوسط	سعودي	مرتفع	غير سعودي

٢٠- إذا كانت $\sum XY = 85, \sum X = 20, \sum Y = 30, \sum X^2 = 165, \sum Y^2 = 200, n = 5$ فأوجد معامل الارتباط للمتغيرين X, Y

٢١- حدد أياً من العبارات التالية صحيح وأيها خطأ مع تصحيح الخطأ:

- يمكن باستخدام شكل الانتشار الحكم على نوع العلاقة بين المتغيرين.

- يفضل استخدام معامل ارتباط بيرسون لتعيين معامل الارتباط بين متغيرين في حالة البيانات الكمية.
- إذا كانت العلاقة بين متغيرين عكسية فإن الارتباط يكون موجباً.
- يستخدم معامل التوافق مع البيانات الوصفية.
- يمكن تعيين التوزيعات الهامشية باستخدام الجداول المزدوجة.
- إذا كان لدينا ظاهرتان لكل منهما صفتان فقط فإنه يتم استخدام معامل الاقتران لتعيين مدى العلاقة بين الظاهرتين.
- يختلف معامل ارتباط بيرسون وسبيرمان للرتب لنفس البيانات من حيث النوع.
- خط الانحدار يستخدم لتعيين مدى العلاقة بين متغيرين.
- معامل ارتباط بيرسون يسمي بمعامل الارتباط اللامعلمي بينما سبيرمان للرتب يسمي بمعامل الارتباط المعلمي.
- يمكن استخدام معادلة الانحدار لتعيين قيم متغير بدلالة المتغير الآخر.

الفصل السادس

الاحتمالات

THE PROBABILITY

٦–١ مقدمة

اقتصرت الإنسانية لفترة طويلة من الزمن أثناء دراستها على ما يدعى بالقوانين الحتمية(Determined). ولكن نظراً لأن جميع الظواهر في حياتنا العملية واليومية تخضع لسنن كونية، ولعل عجز العلماء عن التنبؤ بنتائج بعض هذه المظاهر من خلال قوانين علمية معروفة دفعهم للقول بأن هذه الظواهر تخضع لعامل الصدفة. ومن هنا بدأ التفكير في محاولة تقدير احتمال حدوث حادثة معينة وذلك بعبارات تدل على الصدفة. فنستعمل كلمات مثل " من المحتمل"، "ربما" وغيرها، وعند تفكيرنا بحادثة لم تصبح حقيقة واقعة بعد، أو كانت نتيجتها خارج نطاق سيطرتنا عليها، فإننا نقوم عفوياً بحساب الاحتمال والصدفة.

لقد كانت كلمة احتمال تستعمل بين الأوساط الأرستقراطية في أوروبا في ألعاب الحظ فكانوا يحاولون استشفاف معلومات تساعدهم على الفوز في لعب الورق والنرد. ولم يبدأ حساب الاحتمال بالشكل الذي نعرفه اليوم إلا في منتصف القرن السابع عشر على يد ثلاثة من الفرنسيين هم: دوفرما، باسكال، دوميريه، كما كان مبدأ تكافؤ الفروض معروفاً عند العالم ديموافر.

وقد برز في هذه الفترة علم "رياضيات الاحتمال" وتطور هذا الاختصاص الجديد تطوراً ملموساً حتى أصبح يسيطر على مظاهر عديدة من مظاهر الحياة الحديثة؛ فإلى جانب سيطرته على التأمين ساعد علماء الذرة في فهم

الآثار المتشابكة التي تسجلها الجزئيات الذرية المقذوفة من السيكلوترون على الأفلام، وساعد خبراء الصواريخ في تحديد عوامل الأمان التي يجب أن تُزود بها أجهزة القذائف الباهظة الثمن، وساعد علماء النفس في تقدير ذكاء الأطفال عند إجراء اختبارات الذكاء عليهم وساعد رجال الانتخابات في توقع النتائج قبل حدوثها.

وأصبح لنظرية الاحتمال والمصادفة من الأسس الرياضية العلمية ما يجعلها تطبق على نطاق واسع حيثما انعدم الحكم الصحيح المطلق، وتقدمت دراسة نظرية المصادفة والاحتمال من الوجهة الرياضية تقدماً كبيراً، حتى أصبحنا قادرين على التنبؤ بحدوث بعض الظواهر التي نقول أنها حدثت بالمصادفة، والتي لا نستطيع أن نفسر ظهورها بطريقة أخرى (مثل قذف زهرة النرد). وأصبحنا بفضل هذه الدراسات قادرين على التمييز بين ما يمكن أن يحدث بطريقة المصادفة وما يستحيل حدوثه بهذه الطريقة. إلا أن ألعاب الورق والنرد ليست كل شيء في حساب الاحتمال فهو يشكل في مظهره العملي العنصر الرئيس في علم الإحصاء ويدخل حساب الاحتمال الإحصائي بمجال الأعمال التجارية. فيقدر كمية البضائع التي يجب على المنتج أن يخزنها في مخازنه احتياطاً، ليستطيع تغطية الطلبات الغير متوقعة على منتجاته، كما يكشف لمهندسي الاتصالات عدد الاتصالات التي ينبغي أن تنشأ في أي قسم تليفوني أو شبكة برقية، ويستخدم أيضا في صناعة الأدوية ليكشف عما إذا كانت الآثار الجانبية التي تنتج من استعمال متطوعين لدواء جديد هي أثار ذات قيمة إحصائية أم أنها مجرد صدفة. ويبقى فرق أساسي بين ألعاب الحظ وبين هذه التطبيقات الأخرى التي تفوق الأولى تعقيداً وتفضلها فائدة، وهو أنه يمكن في ألعاب الحظ أن نعد كل النتائج الممكنة التي يمكن أن توزع بها كل أوراق اللعب. وقد يكون هذا التعداد صعباً إلا أنه ممكن دوماً أما عندما نرغب في التنبؤ بتقلبات الحياة، فكثيراً ما يستحيل أن نعرف مسبقاً كل أوراق اللعب المطروحة.

وتعد نظرية الاحتمالات إحدى فروع العلوم الرياضية التي تهتم بالظواهر الكمية التي يعبر عنها بمتغيرات عشوائية تتأثر قيمها بعامل الصدفة. وللدخول في طيات هذه النظرية يلزمنا معرفة بعض المبادئ الأساسية والأسس والنظريات العلمية كمفاتيح أساسية لهذا العلم الذي ينمو بصورة لا حدود لها.

٦-٢ التجربة العشوائية (Random Experiment)

هي التجربة التي نعلم جميع النواتج الممكنة لها، وإن كنا لا نستطيع أن نتنبأ أيٌّ هذه النواتج سيظهر عند إجرائها.

مثال(٦-١)

(أ) تجربة إلقاء قطعة عملة.

(ب) تجربة إلقاء زهرة النرد.

(ج) تجربة قرعة كأس العالم لكرة القدم.

٦-٣ فراغ العينة (Sample Space)

هو مجموعة كل النتائج التي يمكن الحصول عليها من تجربة عشوائية. ونرمز لها بالرمز S.

ويوجد ثلاث أنواع لفراغ العينة هي:

(أ) **فراغ العينة المحدود:** وهو الفراغ الذي يحتوى على عدد محدود من العناصر.

ومثال ذلك فضاء العينة في تجربة إلقاء قطعة نقود(عمله) معدنية مرة واحدة،

$$S = \{H , T\}$$

(ب) **فراغ عينة لانهائي محدود:** وهو الفراغ الذي يحتوى على عدد لانهائي من العناصر ولكنه قابل للعد. بمعنى أن هناك تناظراً بين عناصره وعناصر مجموعة الأعداد الطبيعية أو مجموعة جزئية منها.

ومثال ذلك فراغ العينة لتجربة ألقاء قطعة نقود حتى تظهر الصورة

$$S = \{1, 2, 3, \dots, \infty\}$$

حيث العنصر ٣ يشير إلى أن الصورة لم تظهر في الرمية الأولي، والثانية، ولكنها ظهرت في الرمية الثالثة. كما أن ∞ تشير إلى عدم ظهور الصورة وأن قطعة النقود قد ألقيت عدداً كبيراً جداً من المرات.

(ج) **فراغ عينة لانهائي:** هو الفراغ الذي يحتوى على عدد لانهائي من العناصر.

مثال ذلك تجربة اختيار نقطة داخل دائرة، فإن فراغ العينة لهذه التجربة يتكون من جميع النقاط داخل الدائرة وعدد النقاط داخل الدائرة غير محدود. وعلى ذلك فإن فراغ العينة غير منتهٍ ويكتب بالصورة التالية:

$$S = \{(x , y) : x^2 + y^2 \le a^2 \}$$

حيث x الإحداثي السيني ، y الإحداثي الصادي ، a نصف القطر.

مثال (٦-٢)

تجربة اختيار رقم صحيح موجب أقل من 10. فإن فراغ العينة لهذه التجربة هي

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

مثال (٦-٣)

تجربة إلقاء قطعة عملة معدنية مرتين متتاليتين. فإن النتائج الممكنة لهذه التجربة هي

الرمية الأولى / الرمية الثانية	H	T
H	HH	HT
T	TH	TT

وبالتالي فإن فضاء العينة هو

$$S = \{HH, HT, TH, TT\}$$

مثال (٦-٤)

تجربة إلقاء قطعة عملة معدنية ثلاثة مرات متتالية. فإن النتائج الممكنة لهذه التجربة هي

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

مثال (٦-٥)

تجربة إلقاء حجر نرد مرة واحدة. فإن فضاء العينة لهذه التجربة هو

$$S = \{1, 2, 3, 4, 5, 6\}$$

مثال (٦-٦)

تجربة إلقاء حجر نرد مرتين متتاليتين. فإن فضاء العينة لهذه التجربة هو

$$S=\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),$$
$$(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),$$
$$(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$$

ملاحظات (٦-١)

١- عند التعبير باللفظ عن قطعة العملة المعدنية أو زهرة النرد فيكون كل من قطعة العملة وزهرة النرد متزنة تماماً ومتماثلة، أي تكون فرصة ظهور الصورة في قطعة العملة هي نفسها فرصة ظهور الكتابة. كما أن كل رقم على زهرة النرد له نفس فرصة الظهور.

٢- فراغ العينة الناتج من إلقاء قطعة العملة مرتين متتاليتين هو نفسه فراغ العينة الناتج من إلقاء قطعتي عملة معاً مرة واحدة وكذلك زهرة النرد.

٣- يمكن معرفة عدد نتائج التجربة العشوائية من القاعدة التالية:

$$\text{عدد نتائج التجربة العشوائية} = (\text{عدد الأوجه})^{\text{عدد مرات الرمي}}$$

٦-٤ الحدث (The Event)

الحدث(الحادثة) هو أي مجموعة جزئية من فضاء العينة للتجربة العشوائية.

٦-٤-١ أنواع الحوادث

يمكن تقسيم الحوادث من حيث عدد عناصرها إلى نوعين من الحوادث كالتالي:

١- **الحادثة البسيطة (Simple Event)**: وهي الحادثة التي تحتوى على عنصر واحد فقط من عناصر فضاء العينة للتجربة العشوائية.

٢- **الحادثة المركبة (Compound Event)**: هي الحادثة التي تحتوى على أكثر من عنصر من عناصر فضاء العينة للتجربة العشوائية.

وعموما يمكن تقسيم الحوادث من حيث الظهور (الوقوع) إلى ثلاثة أنواع رئيسية هي:

١- **الحادثة المؤكدة (Sure Event):** وهى الحادثة التي تظهر دائماً عند تكرار إجراء التجربة العشوائية ويرمز لها بالرمز S.

مثال ذلك، في تجربة إلقاء قطعة عملة مرة واحدة، حادثة ظهور صور أو كتابة فهي حادثة مؤكدة.

٢- **الحادثة العشوائية (Random Event):** وهى الحادثة التي قد تظهر وقد لا تظهر عند تكرار إجراء التجربة العشوائية. ويرمز لها بالرمز $A, B, C, \ldots$

ومثال ذلك: في تجربة إلقاء قطعة العملة مرة واحدة، حادثة ظهور صورة، فهي حادثة عشوائية.

٣- **الحادثة المستحيلة (Impossible Event)**: وهى الحادثة التي لا تظهر أبداً عند تكرار إجراء التجربة العشوائية. ويرمز لها بالرمز Φ

ومثال ذلك: في تجربة إلقاء قطعة العملة مرة واحدة، حادثة ظهور الرقم 1، فهي حادثة مستحيلة.

مثال (٦-٧)

في تجربة إلقاء قطعة عملة معدنية ثلاث مرات متتالية. أوجد الحوادث التالية مع تحديد عدد عناصر ونوع كل منها:

(أ) ظهور صورة في الرمية الأولى.

(ب) الحصول على صورة في الرمية الثانية والثالثة.

(ج) ظهور صورة واحدة على الأقل.

(د) ظهور صورة أو كتابة.

(ز) ظهور ثلاثة صور.

(و) ظهور أربعة صور.

(هـ) عدم الحصول على صورة في الرمية الأولى.

الحل:

النتائج الممكنة لتجربة إلقاء قطعة العملة ثلاث مرات هي

$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}, \quad n(S) = 8$

(أ) بفرض أن A حادثة ظهور صورة في الرمية الأولى فإن

$A = \{HHH, HHT, HTH, HTT\}, \quad n(A) = 4$

الحادثة A هي حادثة مركبة وأيضاً حادثة عشوائية.

(ب) بفرض أن B حادثة الحصول على صورة في الرمية الثانية والثالثة فإن

$B = \{HHH, THH\}, \quad n(B) = 2$

الحادثة B حادثة مركبة وأيضاً حادثة عشوائية.

(ج) بفرض أن C حادثة ظهور صورة على الأقل فإن

$C = \{HHH, HHT, HTH, HTT, THH, THT, THH\}, \quad n(C) = 7$

الحادثة C هي حادثة مركبة وأيضاً حادثة عشوائية.

(د) بفرض أن D حادثة ظهور صورة أو كتابة فإن

$D = \{HHH, HHT, HTH, THH, THH, THT, TTH, TTT\}, \quad n(D) = 8$

الحادثة D هي حادثة مركبة وأيضا حادثة مؤكدة.

(ز) بفرض أن E حادثة ظهور ثلاثة صور فإن

$$E = \{HHH\}, \qquad n(E) = 1$$

الحادثة E هي حادثة بسيطة وأيضاً حادثة عشوائية.

(هـ) بفرض أن F حادثة ظهور أربعة صور فإن

$$F = \{\,\} = \Phi, \qquad n(F) = 0$$

الحادثة F هي حادثة مستحيلة.

(و) بفرض أن G حادثة عدم الحصول على صورة فإن

$$G = \{TTT\}, \qquad n(G) = 1$$

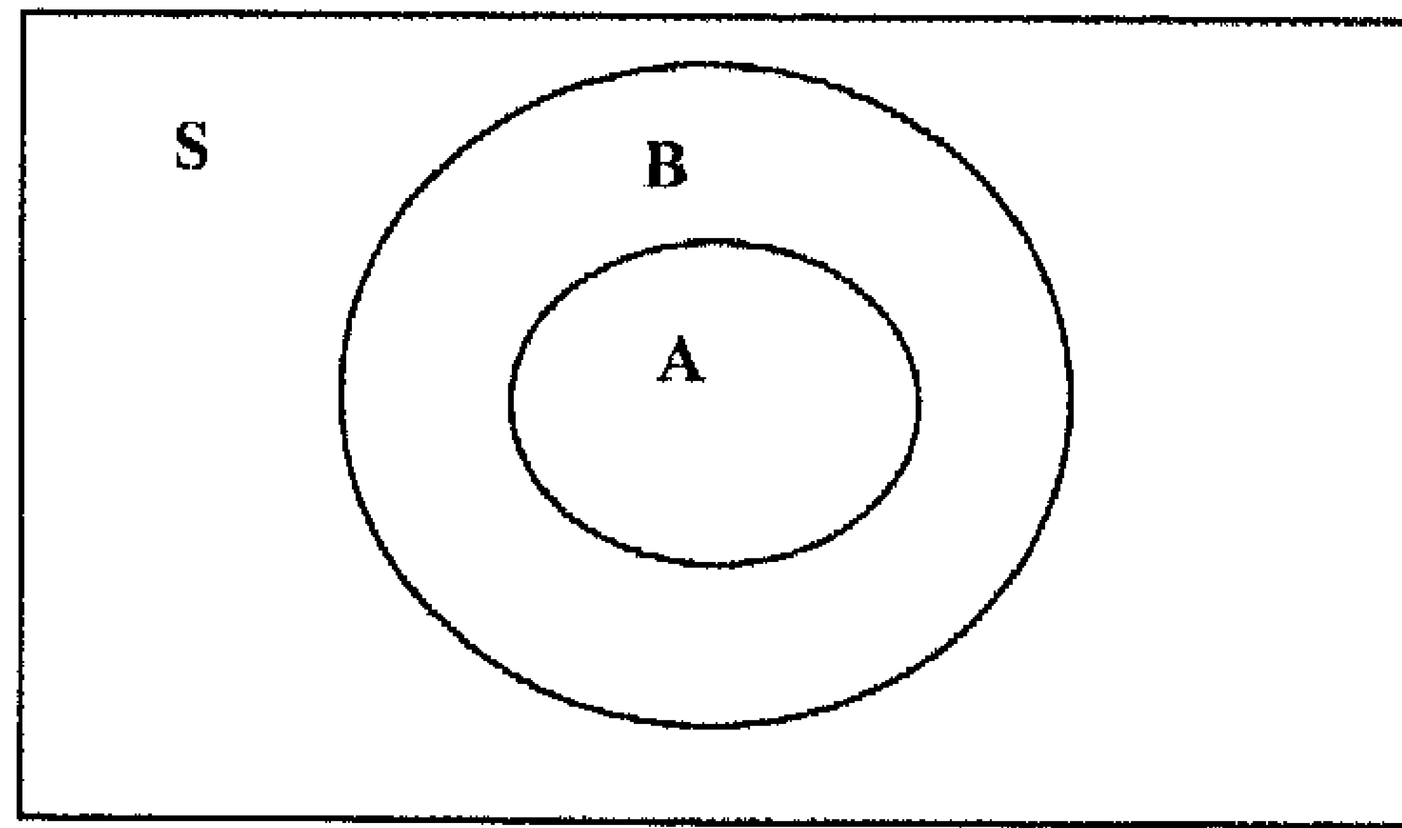
الحادثة G حادثة بسيطة وأيضاً حادثة عشوائية.

٦-٤-٢ العلاقات بين الحوادث (Relation Between Events)

بفرض أن A, B حادثتان معرفتان على نفس فضاء العينة S لتجربة عشوائية ما، فإنه يمكن توضيح العلاقات المختلفة بين الحوادث كالتالي:

١- الحادثة الجزئية (Sub Event)

A حادثة جزئية من B إذا كان كل عناصر A تقع في B ويرمز لها بالرمز $A \subset B$ ويمكن التعبير عنها بشكل فن كالتالي:

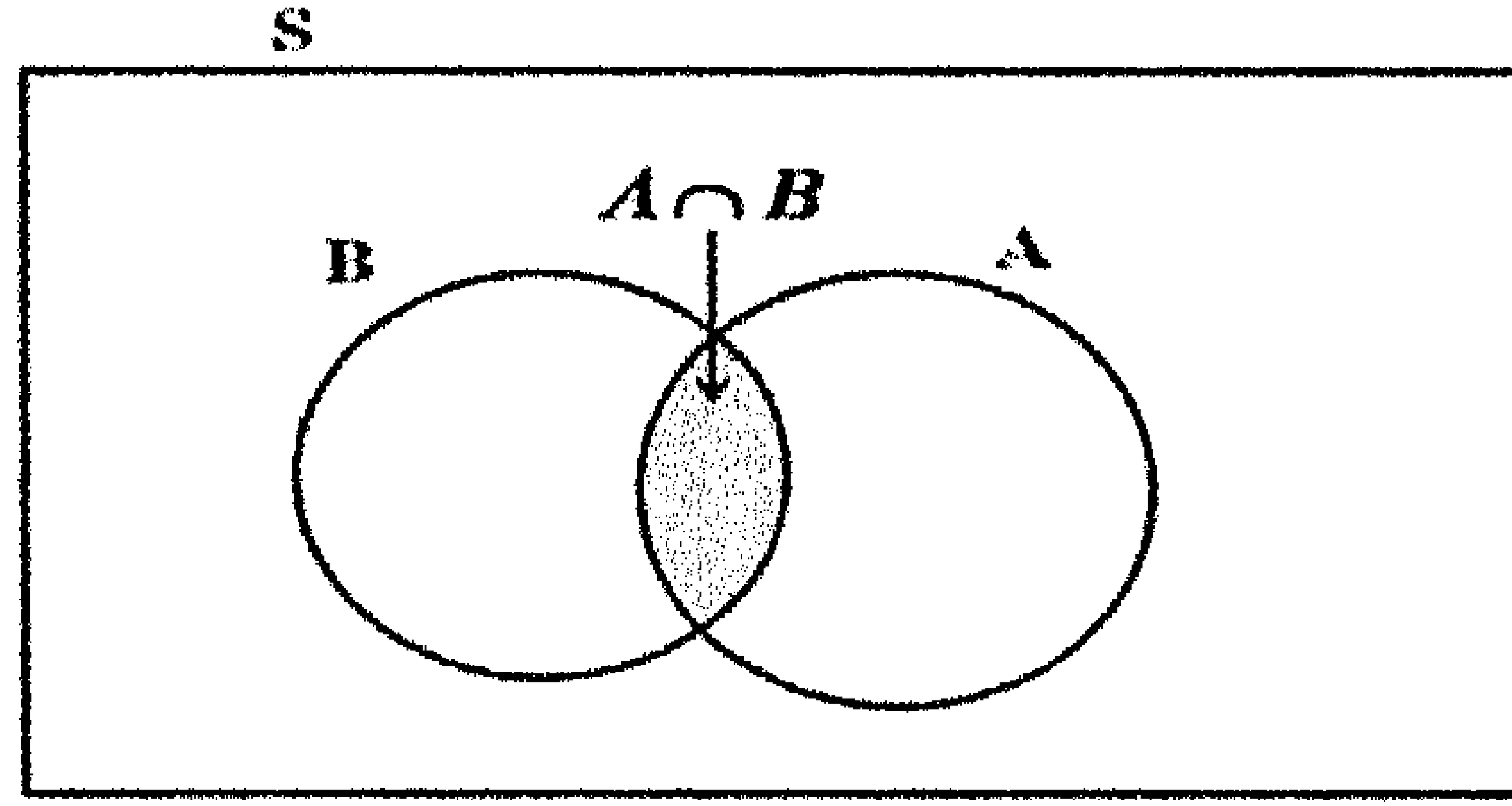


٢- تكافؤ حادثتين (Equivalence of Two Events)

الحادثة A تكافئ الحادثة B وتكتب $A = B$ إذا كانت $A \subset B,\ B \subset A$

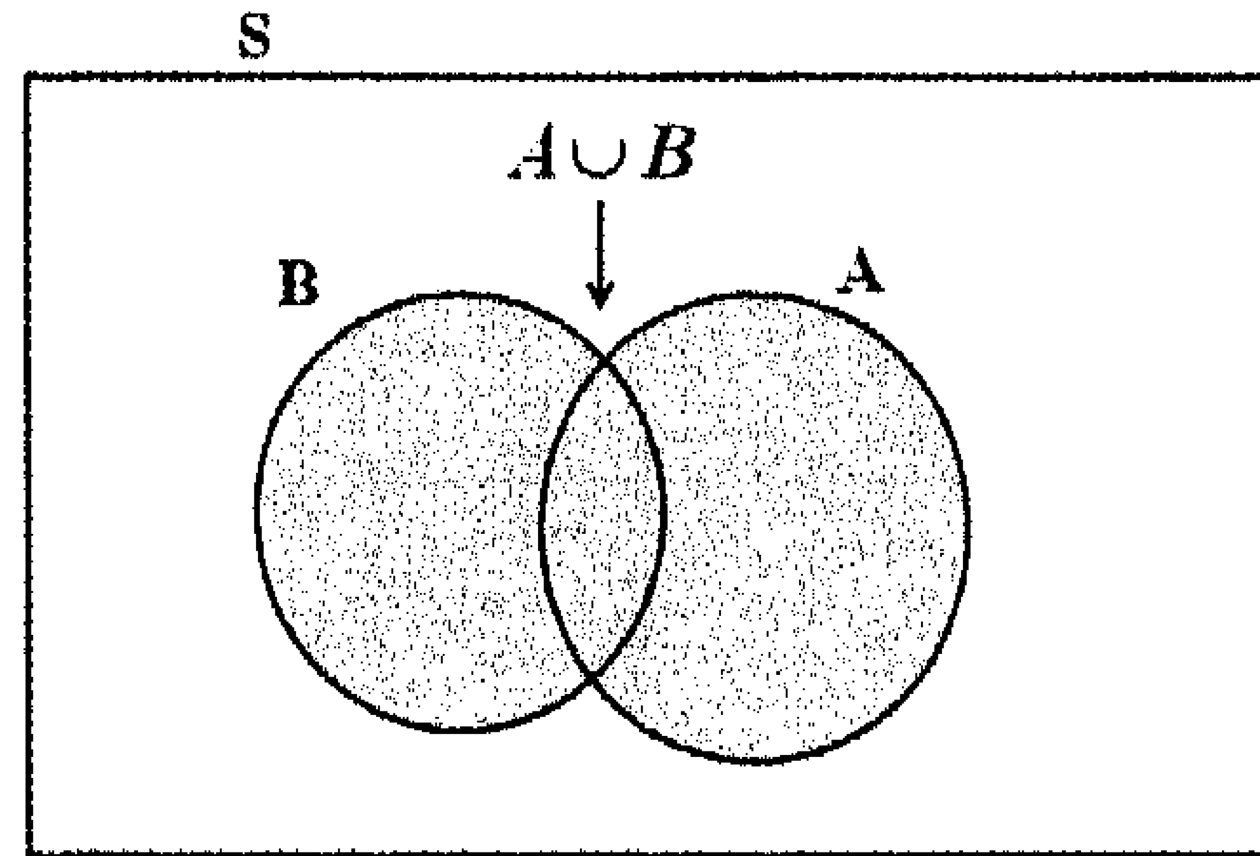
٣- **تقاطع حادثتين (Intersection of Two Event)**

A تقاطع B وهى جميع العناصر التي تقع في A وفي نفس الوقت تقع في B ويرمز لها بالرمز $A \cap B$ ويقع التقاطع إذا وقعت الحادثتان A, B معاً ويمكن التعبير عن التقاطع بشكل فن كالتالي:



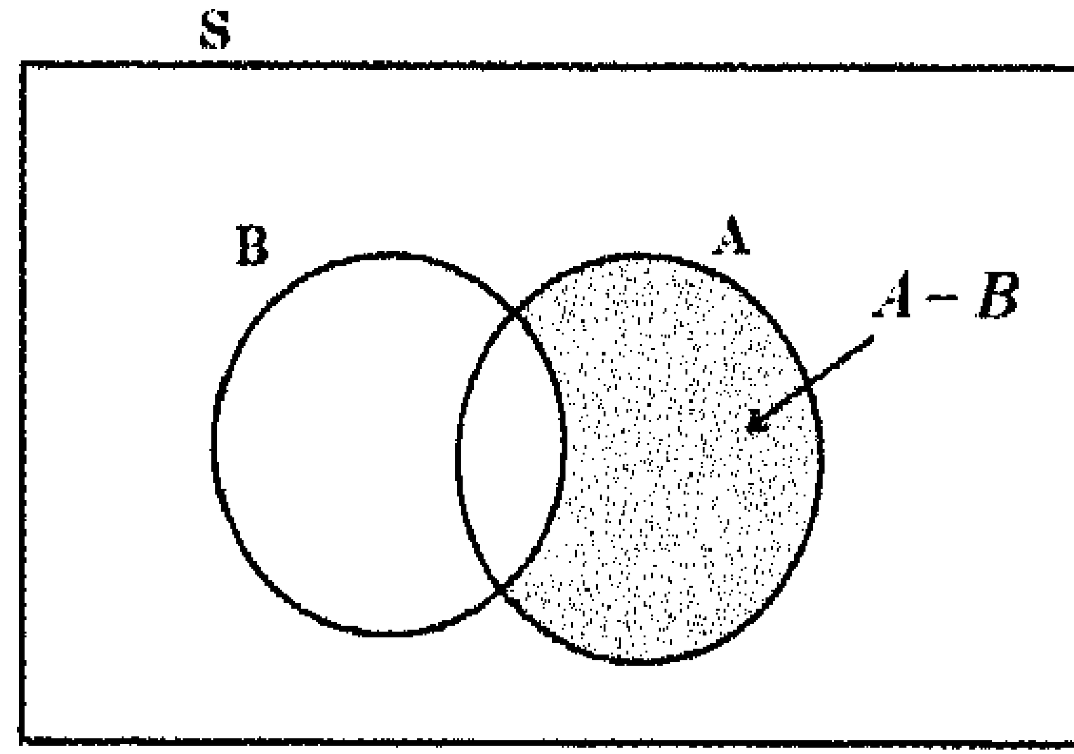
٤- **اتحاد حادثتين (Union of Two Events)**

A اتحاد B وهى جميع العناصر التي تقع في A وجميع العناصر التي تقع في B ويرمز لها بالرمز $A \cup B$ ويقع الاتحاد إذا حدثت A أو B أو كلاهما معاً ويمكن التعبير عن الاتحاد بشكل فن كالتالي:



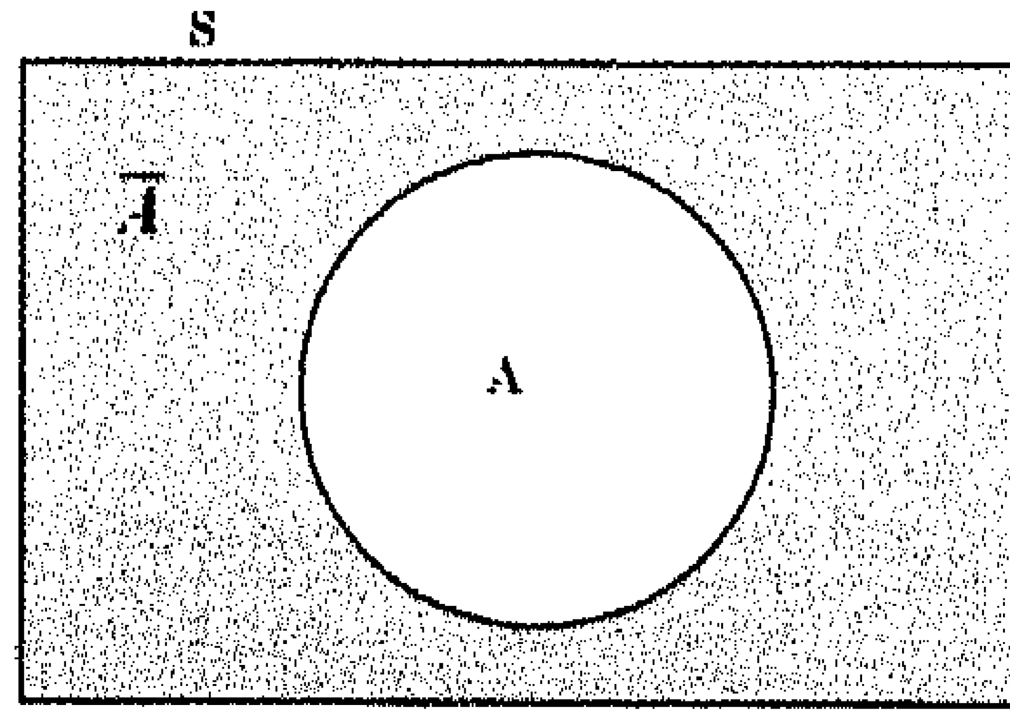
٥- **الفرق بين حادثتين (Difference Between Two Events)**

الفرق بين الحادثتين A, B ويرمز له بالرمز $A - B$ وهى جميع العناصر الموجودة في A وغير موجودة في B ويقع الفرق $A - B$ إذا حدثت A ولم تحدث B ويمكن التعبير عن الفرق بين الحادثتين A, B بشكل فن كالتالي:



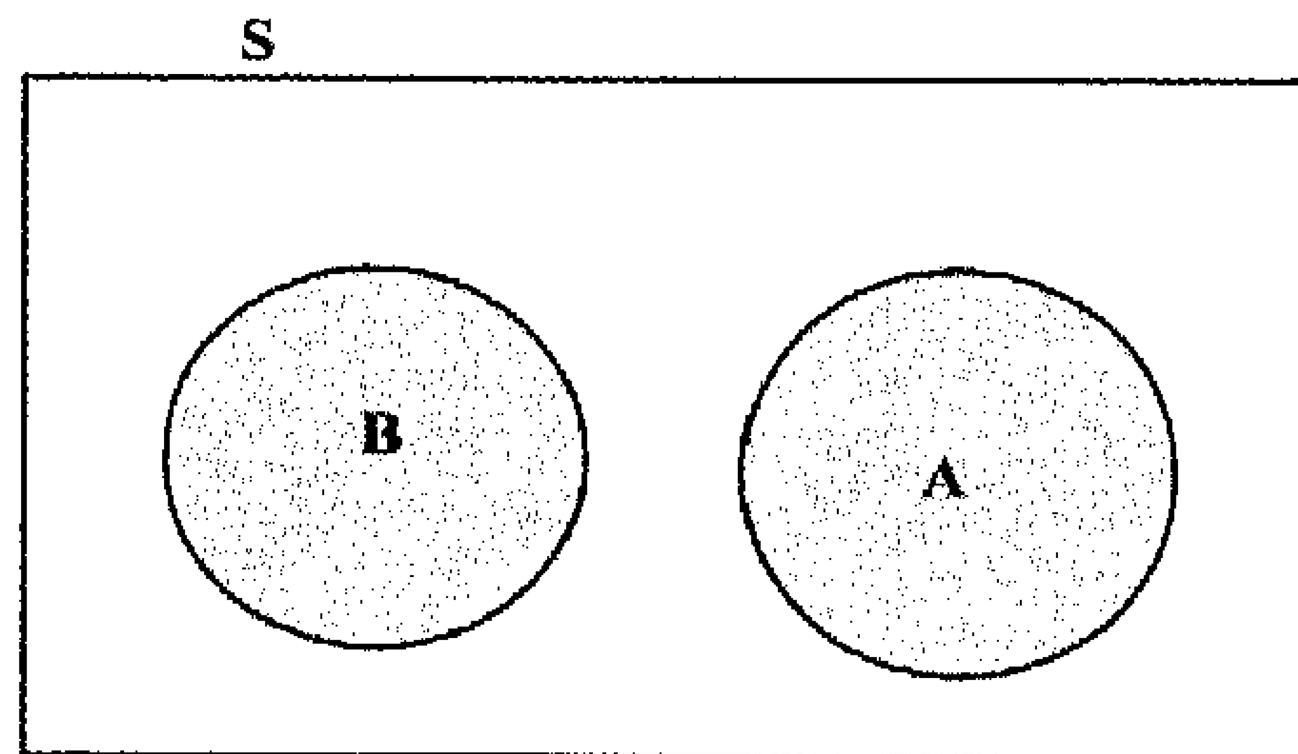
٦- الحادثة المكملة (Complements Event)

الحادثة المكملة للحادثة A هي جميع العناصر التي تقع في فضاء العينة ولا تقع في الحادثة A ويرمز لها بالرمز $\bar{A}$ (أو A^C) وتقع المكملة إذا لم تقع الحادثة A ويمكن التعبير عن مكملة الحادثة A بشكل فن كالتالي:



٧- الحوادث المتنافية (Mutually Exclusive Events)

الحادثتين $A, \ B$ حادثتان متنافيتان إذا كان من المستحيل حدوثهما معاً بمعنى أن $A \cap B = \Phi$ والحوادث المتنافية إذا وقعت إحداهما فلا تقع الأخرى ويمكن التعبير عن الحوادث المتنافية بشكل فن كالتالي:



٨- قوانين دي مورجان: إذا كانت A, B أي حادثتين معرفتين على نفس فضاء العينة لتجربة عشوائية ما، فإن

$$(A \cap B)^C = A^C \cup B^C$$

$$(A \cup B)^C = A^C \cap B^C$$

مثال (٦-٨)

في تجربة إلقاء حجر نرد مرة واحدة، إذا كانت A حادثة ظهور عدد زوجي، B حادثة ظهور عدد فردى، C حادثة ظهور عدد أكبر من 3، فأوجد الحوادث التالية وعدد عناصرها

$$A, B, C, A \cap C, A - C, A \cap B, B \cup C, B - C, A^C, B^C, (A \cup B)^C$$

الحل:

فضاء العينة لهذه التجربة هو

$$S = \{1, 2, 3, 4, 5, 6\}, \qquad n(S) = 6$$

الحوادث

الحادثة	عدد العناصر
$A = \{2, 4, 6\}$,	$n(A) = 3$,
$B = \{1, 3, 5\}$	$n(B) = 3$,
$C = \{4, 5, 6\}$,	$n(C) = 3$,
$A \cap C = \{4, 6\}$,	$n(A \cap C) = 2$,
$A - C = \{2\}$,	$n(A - B) = 1$,
$A \cap B = \{\} = \Phi$,	$n(A \cap B) = 0$,
$B \cup C = \{1, 3, 4, 5, 6\}$,	$n(B \cup C) = 5$,
$B - C = \{1, 3\}$,	$n(B - C) = 2$,
$A^C = \{1, 3, 5\} = B$,	$n(A^C) = 3$,
$B^C = \{2, 4, 6\} = A$,	$n(B^C) = 3$,
$(A \cup B)^C = A^C \cap B^C = \{\} = \Phi$,	$n(A \cup B)^C = 0$

٦-٥ تعريف الاحتمال (Definition of Probability)

بعد التعرض لتعريف الحوادث وأنواعها سوف نتعرض لمفهوم الاحتمال وكيفية تعيين احتمال حدوث الحادثة، فتوجد عدة تعريفات مختلفة للاحتمال، منها:

(أ) التعريف القديم (الكلاسيكي) للاحتمال

(ب) التعريف الرياضي للاحتمال

وفيما يلي سوف نتكلم باختصار عن هذه التعريفات.

أولا التعريف القديم (الكلاسيكي) للاحتمال

إذا كان عدد الطرق التي يمكن أن تظهر بها نتائج تجربة عشوائية ما هو n وكانت هذه النتائج لها نفس فرصة الظهور وكان من بينها m طريقة تظهر بها حادثة ما. فإنه يقال إن احتمال وقوع هذه الحادثة هو $\frac{m}{n}$.
فإذا رمزنا للحادثة بالرمز A وللاحتمال بالرمز $P(.)$ فإن $P(A)$ هو احتمال وقوع الحادثة A وعلى ذلك فإن

$$P(A) = \frac{n(A)}{n(S)} = \frac{m}{n}$$

مثال (٦-٩)

ألقيت زهرة نرد متزنة مرة واحدة. فما هو احتمال ظهور عدد زوجي؟

الحل

فضاء العينة للتجربة هو

$$S = \{1, 2, 3, 4, 5, 6\}$$

بفرض أن A حادثة ظهور عدد زوجي فإن

$$A = \{2, 4, 6\}$$

فإن $n = 6,\ m = 3$ وهذا يؤدى إلى أن احتمال ظهور عدد زوجي (الحادثة A)

$$P(A) = \frac{n(A)}{n(S)} = \frac{3}{6} = 0.5$$

ثانيا التعريف الرياضي للاحتمال

لاحظنا أن التعريف القديم للاحتمالات يشترط تساوي فرص ظهور النتائج الممكنة للتجربة العشوائية، ولكن في بعض التجارب أو المحاولات يكون لبعض النتائج فرص أكبر من غيرها في الظهور. في مثل هذه التجارب لا يمكن استخدام هذا التعريف. ولتفادي ذلك تم وضع التعريف الرياضي للاحتمال وهو مبنى على أساس افتراض بعض المسلمات والتي تسمى مسلمات نظرية الاحتمالات وهى:

المسلمة الأولى: يرافق كل حادثة A عدد معين $P(A)$ يسمى احتمال حدوث(وقوع)الحادثة A ويحقق

$$P(A) \geq 0$$

المسلمة الثانية: احتمال وقوع حادثة مؤكدة يساوى 1 أي أن:

$$P(S) = 1$$

المسلمة الثالثة: إذا كانت A, B حادثتين متنافيتين ومعرفتين على نفس فضاء العينة لتجربة عشوائية ما، فإن

$$P(A \cup B) = P(A) + P(B)$$

والمسلمة الثالثة يمكن تعميمها إلى أكثر من حادثتين، فإذا كانت $A_1, A_2, A_3, \ldots, A_n$ هي n من الحوادث المتنافية بالتبادل فيما بينها والمعرفة على نفس فضاء العينة لتجربة ما، أي أن $A_i \cap A_j = \Phi$ بحيث $i \neq j$ فإن

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

٦-٦ بعض نتائج مسلمات الاحتمالات (Some Results for the Probability Axioms)

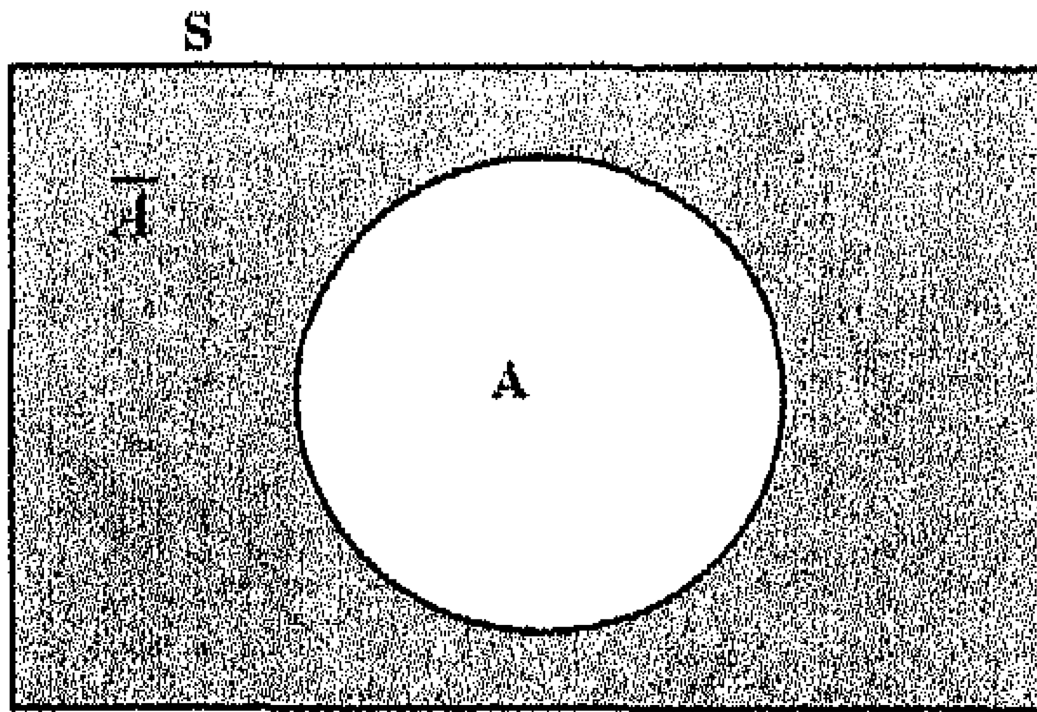
يمكن باستخدام مسلمات الاحتمال السابقة إثبات النتائج التالية:

نتيجة(٦-١)

إذا كانت A^C هي الحادثة المكملة للحادثة A فإن

$$P(A^C) = 1 - P(A)$$

الإثبات



من شكل فن المقابل فإن

$$S = A \cup A^C$$

فإن

$$P(S) = P(A \cup A^C)$$

وحيث إن الحادثتين A, A^C حادثتان متنافيتان فإنه من المسلمة الثالثة للاحتمال فإن

$$P(S) = P(A) + P(A^C)$$

ومن المسلمة الثانية للاحتمال ينتج أن

$$1 = P(A) + P(A^C)$$

$$\Longrightarrow P(A^C) = 1 - P(A)$$

نتيجة(٦-٢)

احتمال وقوع حادثة مستحيلة يساوى صفراً أي أن

$$P(\Phi) = 0$$

الإثبات

نعلم أن

$$S^C = \Phi$$

وأن

$$P(S^C) = P(\Phi)$$

ولكن من نتيجة (٦-١) فإن

$$P(S^C) = 1 - P(S)$$

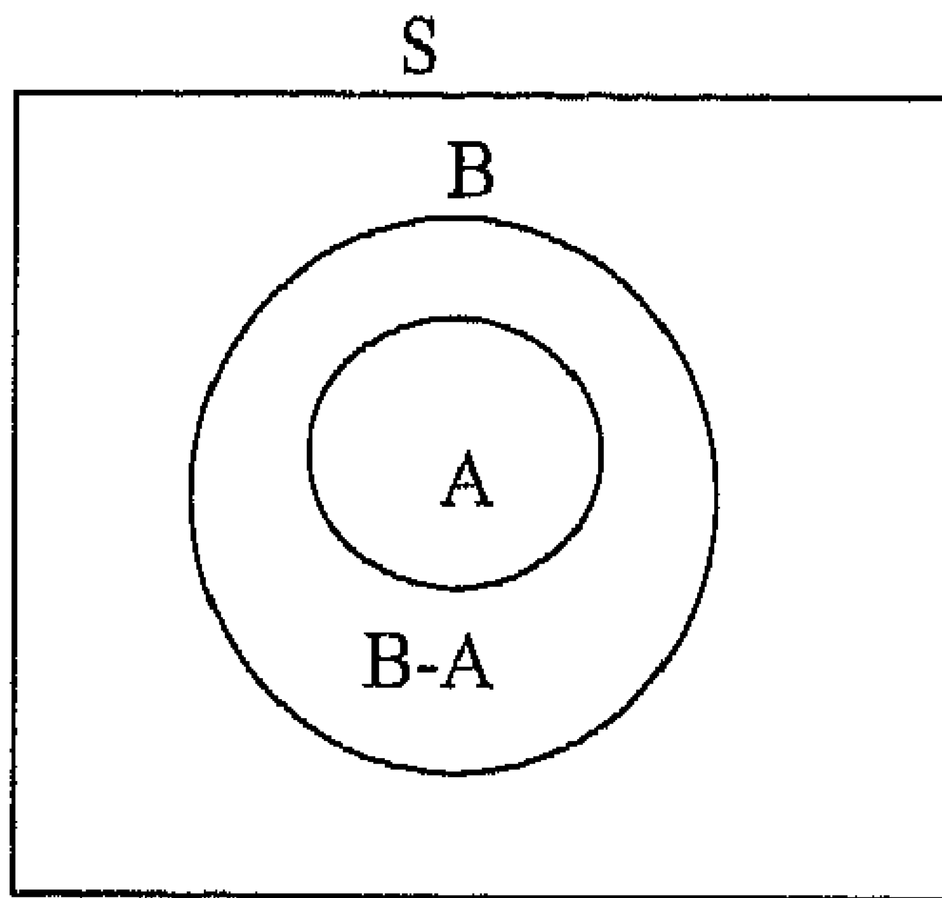
ومن المسلمة الثانية فإن

$$P(S) = 1$$

$$\therefore P(\Phi) = P(S^C) = 1 - 1 = 0$$

نظرية (٦-١)

إذا كانت $A \subseteq B$ فإن $P(A) \leq P(B)$



الإثبات

من شكل فن المقابل وحيث أن $A \subseteq B$

فإن

$$B = A \cup (B - A)$$

$$P(B) = P[A \cup (B - A)]$$

وحيث إن الحادثتين $A, B - A$ متنافيتان

فإنه من المسلمة الثالثة للاحتمال ينتج أن

$$P(B) = P(A) + P(B - A)$$

$$P(B) - P(A) = P(B - A)$$

ولكن من المسلمة الأولى للاحتمال فإن

$$P(B - A) \geq 0$$

فإن

$$P(B) - P(A) \geq 0$$

$$\therefore P(A) \leq P(B)$$

نتيجة(٦-٣)

إذا كانت $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n$ فإن

$$P(A_1) \leq P(A_2) \leq \cdots \leq P(A_n)$$

نتيجة(٦-٤)

لأي حادثة A فإن $P(A) \leq 1$

الإثبات

نعلم أن $A \subseteq S$ ومن النظرية (٦-١) فإن

$$P(A) \leq P(S)$$

لكن من المسلمة الثانية للاحتمال فإن

$$P(S) = 1$$

فينتج أن

$$P(A) \leq 1$$

يلاحظ أنه من المسلمة الأولى والنتيجة (٦-٤) نستنتج أنه لأي حادثة A فإن

$$0 \leq P(A) \leq 1$$

نظرية (٦-٢)

إذا كانت $A,\ B$ أي حادثتين معرفتين على نفس فضاء العينة لتجربة ما، فإن

$$P(B - A) = P(B) - P(A \cap B)$$

الإثبات

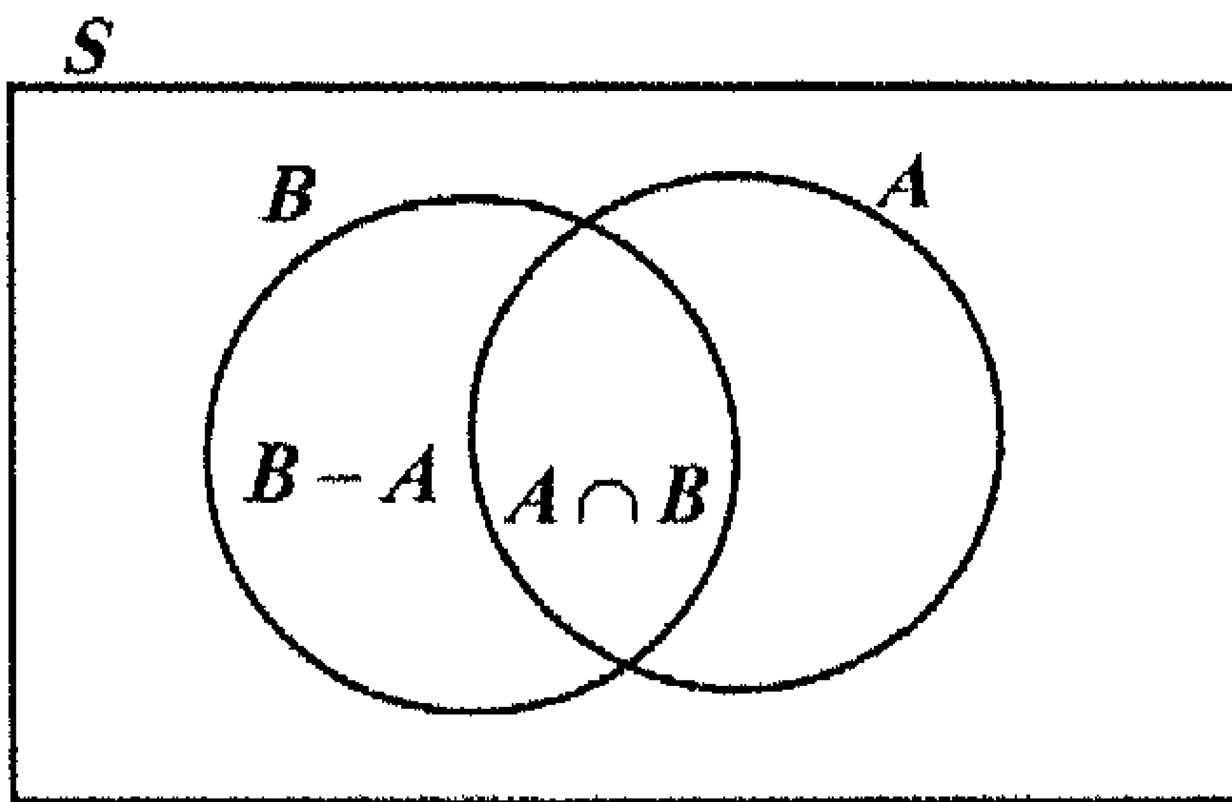
من شكل فن نلاحظ أن

$$B = (B - A) \cup (A \cap B)$$

وكذلك الحادثتان $B - A, A \cap B$ متنافيتان فمن المسلمة الثالثة للاحتمال ينتج أن

$$P(B) = P(B - A) + P(A \cap B)$$

$$\Longrightarrow P(B - A) = P(B) - P(A \cap B)$$



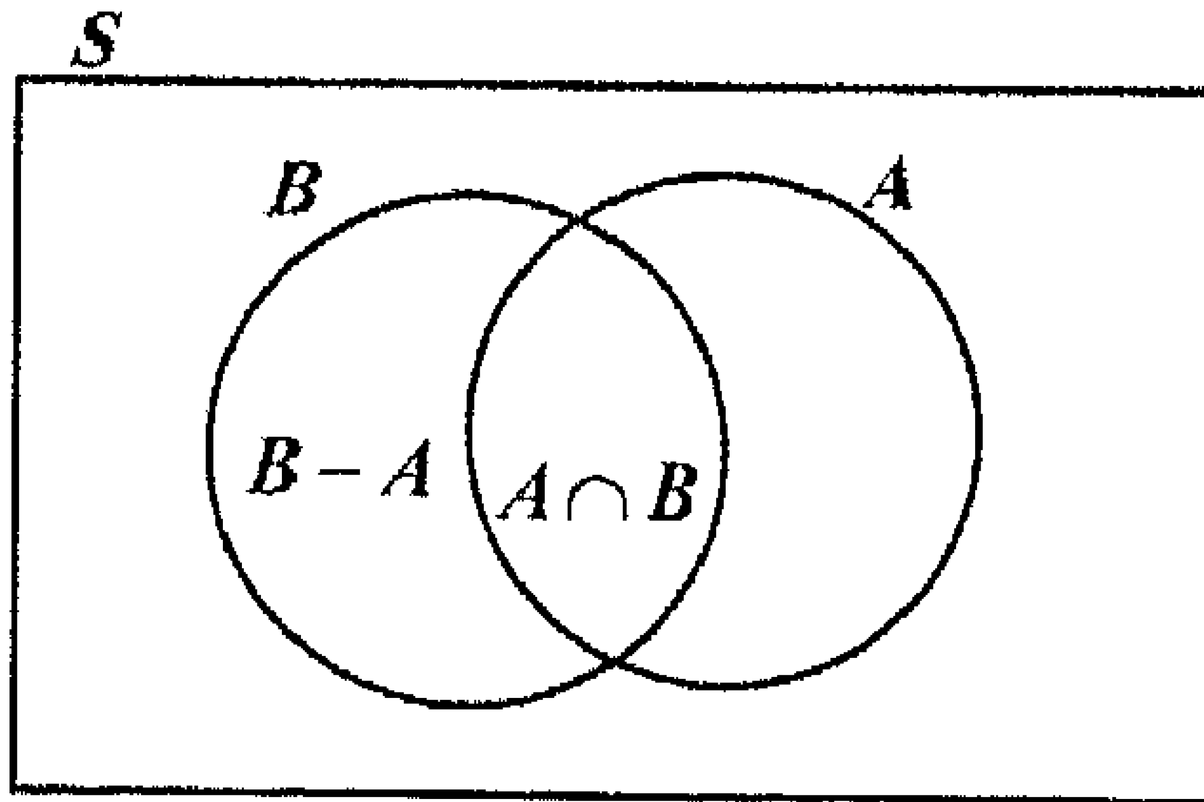
نظرية (٦-٣)

إذا كانت A, B أي حادثتين فإن

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

البرهان

من شكل فن نلاحظ أن



$$A \cup B = A \cup (B - A)$$

وكذلك الحادثتان $B - A$ ، A متنافيتان فمن المسلمة الثالثة للاحتمال ينتج أن

$$P(A \cup B) = P(A) + P(B - A)$$

وباستخدام النظرية (٦-٢) والعلاقة السابقة ينتج أن

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

مثال(٦-١٠)

إذا كان احتمال رسوب طالب في مقرر 1080 إحص هو 0.2 واحتمال نجاحه في مقرر 1080ريض هو 0.7 واحتمال نجاحه في مقرر 1080 إحص ورسوبه في مقرر 1080ريض هو 0.08 فأوجد احتمال

(أ) نجاح الطالب في مقرر 1080 إحص.

(ب) نجاح الطالب في كلٍّ من المقررين.

(ج) نجاح الطالب في أحد المقررين على الأقل.

(د) رسوب الطالب في كلٍّ من المقررين.

الحل:

بفرض إن

A حادثة نجاح الطالب في مقرر 1080 إحص، B حادثة نجاح الطالب في مقرر 1080 ريض

من المعطيات فإن

$$P(A^C) = 0.2, \qquad P(B) = 0.7, \qquad P(A \cap B^C) = 0.08$$

(أ) نجاح الطالب في مقرر 1080 إحص

$$P(A) = 1 - P(A^C) = 1 - 0.2 = 0.8$$

(ب) نجاح الطالب في كلٍّ من المقررين $(A \cap B)$

$$\because \; P(A \cap B^C) = P(A) - P(A \cap B)$$

$$\Longrightarrow 0.08 = 0.8 - P(A \cap B)$$

$$\Longrightarrow P(A \cap B) = 0.8 - 0.08 = 0.72$$

(ج) نجاح الطالب في أحد المقررين على الأقل $(A \cup B)$

$$\because P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= 0.8 + 0.7 - 0.72 = 0.78$$

(د) رسوب الطالب في كلٍّ من المقررين $(A^C \cap B^C)$

من قانون دي مورجان فإن

$$(A \cup B)^C = A^C \cap B^C$$

$$\Longrightarrow P(A^C \cap B^C) = P(A \cup B)^C$$

$$= 1 - P(A \cup B) = 1 - 0.78 = 0.22$$

مثال(٦-١١)

في إطار النشاط الرياضي قررت كلية العلوم بجامعة الخرج إجراء مسابقة رياضية بين ثلاثة فرق رياضية، الفريق الأول A يمثل أعضاء هيئة التدريس بالكلية، والفريق الثاني B يمثل الموظفين الإداريين بالكلية، والفريق الثالث C يمثل طلاب الكلية. فإذا كانت فرصة فوز الفريق B ضعف فرصة فوز الفريق A وكانت فرصة فوز الفريق C ضعف فرصة فوز الفرقة B فأوجد احتمال فوز كل فريق على حده.

الحل

من المعطيات فإن:

$$P(C) = 2P(B)$$

$$P(B) = 2\,P(A)$$

وفى نهاية المسابقة سيفوز فريق واحد فقط. لذا فإن

$$P(A \cup B \cup C) = 1$$

الحوادث A, B, C هي حوادث متنافية وبالتالي من مسلمة الاحتمال الثالثة فإن

$$P(A) + P(B) + P(C) = 1$$

$$\Rightarrow P(A) + 2\,P(A) + 4\,P(A) = 1$$

$$\Rightarrow (1 + 2 + 4)\,P(A) = 1$$

$$\Rightarrow 7P(A) = 1 \;\Rightarrow P(A) = \frac{1}{7}$$

ومما سبق فإن

$$P(A) = \frac{1}{7},\;\; P(B) = \frac{2}{7},\qquad P(C) = \frac{4}{7}$$

٦–٧ طرق العد (Counting Methods)

من التعريف الكلاسيكي للاحتمال نجد أنه إذا كانت نتائج التجربة العشوائية متساوية الاحتمال، فإن طريقة حساب احتمال أي حادثة تتطلب معرفة عددين:

(أ) عدد النتائج الممكنة للتجربة العشوائية أي عدد عناصر فضاء العينة.

(ب) عدد عناصر الحادثة المعنية بالدراسة.

ونحصل على هذه الأعداد عن طريق حصر النتائج الممكنة لفراغ العينة وعدها بالطريقة العادية. ولكن اتضح أنه في بعض المسائل تكون عملية الحصر غير ممكنة كما يظهر ذلك في الأمثلة الآتية.

مثال (٦–١٢)

تم ترتيب ثلاثة أشخاص a, b, c عشوائياً في طابور للحصول على خدمة معينة. احسب احتمال أن يكون الشخص a بجوار الشخص b.

الحل

من السهل حصر فضاء العينة لهذه التجربة وهو

$$S = \{abc, acb, bac, bca, cab, cba\}$$

نفرض أن A حادثة أن الشخص a بجوار الشخص b

$$A = \{abc, bac, cab, cba\}$$

$$P(A) = \frac{n(A)}{n(S)} = \frac{4}{6} = \frac{2}{3}$$

في المثال السابق إذا كان عدد الأشخاص خمسة a, b, c, d, e فما هو احتمال أن يكون الشخص a بجوار b؟ سنلاحظ أن عملية العد أصبحت مشكلة كبيرة وسوف نرى فيما بعد أن فضاء العينة سيكون مكوناً من 120 طريقة سوف ينتج منها 48 طريقة فيها a بجوار b. لكن إذا كان هناك 10 أشخاص بدلاً من 5 فإننا سوف نجد أن هناك 3628800 طريقة يحتويها فضاء العينة منها 725760 طريقة يكون فيها a بجوار b.

والآن سوف نبحث عن مبدأ يساعدنا في إيجاد عدد الطرق الممكنة لترتيب n من الأشياء في صف بدون اللجوء إلى عملية حصر النتائج وعدها. ولتقديم هذا المبدأ نأخذ المثال التالي.

مثال (٦-١٣)

إذا كان لدينا 4 كتب a, b, c, d فبكم طريقة يمكن ترتيبها على رف؟

الحل

إن هذه المشكلة تتمثل في وضع تخصيص 4 أماكن لهذه الكتب، ونلاحظ أن:

١- المكان الأول يمكن وضع أي من الأربعة كتب به، لذا يوجد 4 طرق للمكان الأول.

٢- المكان الثاني يمكن وضع أي من الثلاثة كتب المتبقية به، لذا يوجد 3 طرق للمكان الثاني.

٣- المكان الثالث يمكن وضع أي من الكتابين المتبقيين به، لذا يوجد طريقتان للمكان الثالث.

٤- المكان الرابع سوف يوضع به الكتاب المتبقي. لذا يوجد طريقة واحدة للمكان الرابع.

ويمكن تمثل ما سبق بالتالي:

الرابع	الثالث	الثاني	الأول	الأماكن
1	2	3	4	الطرق

وبذلك يكون عدد الطرق الكلى لوضع 4 كتب على رف هو

$$4 \times 3 \times 2 \times 1 = 24$$

وهذه الطرق هي

abcd	abdc	acbd	acdb	adbc	adcb
bacd	badc	bcad	bcda	bdac	bdca
cabd	cadb	cbad	cbda	cdab	cdba
dabc	dacb	dbac	dbca	dcab	dcba

ومن المثال السابق تتضح لنا نقطتان مهمتان هما:-

١- مبدأ العد.

٢- التباديل.

٦-٧-١ مبدأ العد (Counting Technique)

إذا أمكن إجراء عملية ما بطرق عددها m (وبعد إتمامها) أمكن إجراء عملية أخرى بطرق عددها n فإنه يمكن إجراء العمليتين معاً بطرق عددها $m\,n$.

مثال (٦-١٤)

بكم طريقة يمكن تكوين عدد مكون من رقمين من الأرقام 3 ,4 ,5 بحيث:

(أ) يسمح بتكرار رقم

(ب) لا يسمح بتكرار رقم

الحل

(أ) في حالة السماح بتكرار الرقم

نعلم أن العدد المكون من رقمين يحتوى على خانتين الآحاد والعشرات، ونتيجة للسماح بالتكرار فإن أياً من الأرقام الثلاثة يمكنه الظهور في أي من الخانتين أو الاثنتين معاً.

عدد طرق اختيار الخانة الأولى = 3

عدد طرق اختيار الخانة الثانية = 3

عدد طرق اختيار العدد = $3 \times 3 = 9$

وهذه الطرق(الأعداد) هي

55	54	53
45	44	43
35	34	33

(ب) في حالة عدم السماح بتكرار الرقم، فإن الأرقام التي ستظهر في خانة الآحاد أو العشرات لن تظهر في الخانة الثانية في نفس الوقت وبالتالي فإن:

عدد طرق اختيار الخانة الأولى = 3

عدد طرق اختيار الخانة الثانية = 2

عدد طرق اختيار العدد = 2×3 = 6

وهذه الطرق هي

53	54
43	45
34	35

تعميم مبدأ العد

إذا أمكن إجراء عملية ما بطرق عددها n_1 وأمكن إجراء عملية ثانية بطرق عددها n_2،....... وأمكن إجراء عملية k بطرق عددها n_k بحيث لا يمكن إجراء أي عملية إلا بعد الانتهاء من العملية السابقة لها. فإنه يمكن إجراء هذه العمليات معاً بطرق عددها

$$n_1 \times n_2 \times \ldots \times n_k$$

مثال (٦-١٥)

يقدم مطعم 4 أصناف من اللحم، 3 أصناف من السلطة، 5 أصناف من الحلوى. كم عدد الوجبات المختلفة التي يمكن أن يقدمها. (علماً بأن الوجبة تتكون من لحم وسلطة وحلوى).

الحل

بما أن الوجبة يجب أن تتكون من لحم وسلطة وحلوى فإنه يجب اختيار نوع من اللحم ونوع من السلطة ونوع من الحلوى معاً لذا فإن:

عدد طرق اختيار (تقديم) اللحم = 4

عدد طرق اختيار (تقديم) السلطة = 3

عدد طرق اختيار (تقديم) الحلوى = 5

وبالتالي فإن عدد الوجبات المختلفة التي يمكن تقديمها = 5×3×4 = 60

٦-٧-٢ التباديل (Permutation)

إن تبديل مجموعة من الأشياء يعني تنظيمهما في ترتيب محدد. أي أن تبديل عدد من الأشياء هو تنظيم هذه الأشياء في ترتيب محدد.

نظرية (٦-٤)

عدد تباديل n من العناصر (الأشياء) المختلفة، مأخوذ منها r من العناصر يرمز له بالرمز $^{n}P_r$ ويعطى كما يلي:

$$^{n}P_r = \frac{n!}{(n-r)!} = n(n-1)(n-2)\ldots(n-r+1)$$

حيث $n!$ مضروب العدد الصحيح n ويعطى كالتالي

$$n! = n(n-1)(n-2)\ldots(3)(2)1$$

وأيضاً من خواص التباديل فإن:

$$^{n}P_0 = 1, \qquad ^{n}P_1 = n, \qquad ^{n}P_n = n!\,, \;\; n! = n\,(n-1)!$$

مثال (٦-١٦)

بكم طريقة يمكن صف عشرة أشخاص في صف.

الحل

عند صف مجموعة من الأشخاص في صف فإنه يوجد لدينا عدد من الأماكن مساوٍ لعدد هؤلاء الأشخاص. وبذلك فإن $n = r = 10$

عدد الطرق= $^{n}P_r = \; ^{10}P_{10} = 10! = 3628800$

مثال (٦-١٧)

بكم طريقة يمكن ترتيب أو تبديل سبعة كتب على رف؟

الحل

من الواضح أنه لترتيب الكتب على رف فإن كل كتاب سوف يوضع في مكان واحد فقط وبالتالي سوف يكون لدينا عدد من الأماكن مساوٍ لعدد الكتب الموجودة ومنها فإن $n = r = 7$ وعدد الطرق هي:

$$\text{عدد الطرق} = 7! = 5040$$

مثال (٦-١٨)

بكم طريقة يمكن أن نجلس شخصين على مقعدين من مجموعة مكونة من عشرة أشخاص.

الحل

من الملاحظ انه يوجد لدينا عشرة أشخاص ونريد ترتيب اثنين منهم في مكانين (مقعدين) فقط وبالتالي فإن
$n = 10, r = 2$

$$\text{عدد الطرق} = \; {}^{n}P_r = \; {}^{10}P_2 = 10 \times 9 = 90$$

نظرية (٦-٤)

إذا كان لدينا n من العناصر منها n_1 من العناصر المتشابهة معا، n_2 من العناصر المتشابهة معاً والمختلفة عن باقي العناصر، ....، n_k من العناصر المتشابهة معاً والمختلفة عن باقي العناصر بحيث

$$n_1 + n_2 + \cdots + n_k = n$$

فإن عدد التباديل الممكنة من n هي

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1!\, n_2! \dots n_k!}$$

مثال(٦-١٩)

أوجد عدد تباديل كلمة Statistics.

الحل

نجد أن كلمة Statistics توجد بها 10 حروف، فيها بعض الحروف المكررة لذا يمكن تقسيم الكلمة كالتالي:

الحرف	s	t	a	i	c
عدد مرات التكرار	3	3	1	2	1

$$\text{عدد الطرق} = \binom{n}{n_1, n_2, n_3, n_4, n_5}$$

$$= \binom{10}{3, 3, 1, 2, 1} = \frac{10!}{3!\,3!\,1!\;2!\,1!} = 50400$$

٦-٧-٣ التوافيق (Combinations)

التوفيقة كل مجموعة يمكن اختيارها من مجموعة من عدة أشياء بأخذها كلها أو بعضها دون مراعاة الترتيب.

نظرية (٦-٥)

إذا كان لدينا مجموعة مكونة من n من العناصر المختلفة فإن عدد التوافيق التي يمكن تكوينها بحيث تحوي كل توفيقه على r عنصر يرمز له بالرمز $\binom{n}{r}$ أو بالرمز nC_r ويعطي بالصيغة التالية:

$$\binom{n}{r} = nC_r = \frac{n!}{r!\,(n-r)!}, \qquad r = 0, 1, \dots, n$$

وأيضاً من خواص التوافيق فإن:

$$\binom{n}{0} = 1, \qquad \binom{n}{1} = n, \quad \binom{n}{n} = 1$$

مثال(٦-٢٠)

أوجد قيمة كلٍّ من

$$\binom{5}{3}, \ \binom{4}{0}, \ \binom{7}{7}, \ \binom{7}{1}$$

الحل

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = \frac{5!}{3!\,2!} = 10,$$

$$\binom{4}{0} = \frac{4!}{0!\,(4-0)!} = \frac{4!}{1 \times 4!} = 1,$$

$$\binom{7}{7} = \frac{7!}{7!\,(7-7)!} = \frac{7!}{7!\,0!} = 1,$$

$$\binom{7}{1} = \frac{7!}{1!\,(7-1)!} = \frac{7!}{1!\,6!} = 7$$

مثال(٦-٢١)

بكم طريقة يمكن اختيار حرفين(بدون مراعاة الترتيب) من مجموعة الحروف A, B, C؟

الحل

من الملاحظ أن $n = 3,\ r = 2$ وعليه فإن عدد طرق اختيار حرفين من مجموعة الحروف A, B, C يساوي:

$$\binom{n}{r} = \binom{3}{2} = \frac{3!}{2!\,(3-2)!} = \frac{3!}{2!\,1!} = 3$$

٦-٨ المعاينة — عدد العينات

من الظواهر العشوائية الأساسية التي نهتم بدراستها في نظرية الاحتمالات هي المعاينة. وبفرض أن لدينا وعاء به n من العناصر ونريد سحب عينة حجمها r من تلك العناصر، فيجب معرفة نوع العينة التي سنقوم بسحبها هل سنهتم فيها بالترتيب فتسمى عينات مرتبة أم أنها عينة غير مرتبة، أي ليس هناك اهتمام بترتيب العناصر بها.

٦-٨-١ العينات المرتبة (Ordered Sample)

والغرض هنا هو حساب عدد الطرق التي يمكن بها سحب عينات مرتبة حجمها r من مجموعة من العناصر عددها n. وعملية سحب العينات المرتبة يمكن أن تتم بطريقتين هما:

(أ) السحب بإرجاع (Sampling with Replacement)

يقال إن العينة تم سحبها بإرجاع إذا كان بعد كل عملية سحب(اختيار) نقوم بتسجيل العنصر المسحوب ثم يعاد قبل عملية السحب الثانية، وهكذا تكرر عملية السحب والتسجيل ثم الإعادة قبل عملية السحب التي تليها حتى نحصل على العدد المطلوب من العناصر، ونلاحظ أنه في كل مرة من مرات السحب يكون عدد العناصر ثابتاً. وتكون عدد الطرق اللازمة لسحب عينة حجمها r من n من العناصر بإرجاع هي:

$$\underbrace{n \times n \times n \times \ldots \times n}_{r\ \text{من المرات}} = n^r$$

(ب) السحب بدون إرجاع (Sampling without Replacement)

يقال إن العينة تم سحبها بدون إرجاع إذا كان بعد إجراء كل عملية سحب نقوم بتسجيل العنصر الذي تم سحبه ثم لا يعاد هذا العنصر قبل عملية السحب الثانية وهكذا تكرر عملية السحب وعدم الإعادة قبل عملية السحب التي تليها حتى نحصل على العدد المطلوب من العناصر. ونلاحظ أنه في كل مرة من مرات السحب يقل

عدد العناصر التي سيتم السحب منها بمقدار عنصر عن المرة السابقة لها. وتكون عدد الطرق اللازمة لسحب عينة حجمها r من n من العناصر بدون إرجاع هي:

$$\underbrace{n(n-1)(n-2)\dots(n-r+1)}_{r \text{ من المرات}} = \ ^{n}P_{r}$$

مثال (٦-٢٢)

سحبت كرتان بإرجاع من صندوق يحتوى 4 كرات بيضاء، كرتين حمراء. أوجد احتمال أن تكون

(أ) كل من الكرتين بيضاء.

(ب) كل من الكرتين لهما نفس اللون.

(ج) على الأقل واحدة من الكرتين ستكون بيضاء.

الحل

بفرض أن A حادثة أن كلاً من الكرتين بيضاء، B حادثة أن كلاً من الكرتين حمراء.

فعدد الطرق اللازمة لسحب كرتين بإرجاع من جميع الكرات الموجودة في الصندوق هي عدد عناصر فضاء العينة ويمكن تعيينها كالتالي:

$$n(S) = 6^2 = 36$$

(أ) احتمال أن كلاً من الكرتين بيضاء $P(A)$

عدد الطرق اللازمة لسحب كرتين بإرجاع من الكرات البيضاء

$$n(A) = 4^2 = 16,$$

$$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{16}{36}.$$

(ب) احتمال أن كلاً من الكرتين لهما نفس اللون $P(A \cup B)$

$$P(A \cup B) = P(A) + P(B)$$

عدد الطرق اللازمة لسحب كرتين بإرجاع من الكرات الحمراء

$$n(B) = 2^2 = 4,$$

$$\therefore P(B) = \frac{n(B)}{n(S)} = \frac{4}{36},$$

$$\Rightarrow P(A \cup B) = \frac{16}{36} + \frac{4}{36} = \frac{20}{36}.$$

(ج) احتمال أنه على الأقل كرة واحدة من الكرتين ستكون بيضاء. هذه الحادثة تكافؤ أن لا تكون الكرتين معا لهما اللون الأحمر وبالتالي يكون المطلوب تعيين $P(B^c)$

$$P(B^c) = 1 - P(B) = 1 - \frac{4}{36} = \frac{32}{36}.$$

مثال (٦-٢٣)

في المثال السابق إذا كان السحب بدون إرجاع. احسب قيم الاحتمالات السابقة.

الحل

باستخدام نفس الرموز السابقة، فإن عدد الطرق اللازمة لسحب كرتين بدون إرجاع من جميع الكرات الموجودة في الصندوق هي

$$n(S) = \ {}^6P_2 = 6 \times 5 = 30$$

(أ) احتمال أن كلاً من الكرتين بيضاء $P(A)$

عدد الطرق اللازمة لسحب كرتين بدون إرجاع من الكرات البيضاء

$$n(A) = \ {}^4P_2 = 4 \times 3 = 12,$$

$$\therefore P(A) = \frac{n(A)}{n(S)} = \frac{12}{30}.$$

(ب) احتمال أن كلاً من الكرتين لهما نفس اللون $P(A \cup B)$

$$P(A \cup B) = P(A) + P(B)$$

عدد الطرق اللازمة لسحب كرتين بدون إرجاع من الكرات الحمراء

$$n(B) = 2P_2 = 2 \times 1 = 2,$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{2}{30},$$

$$P(A \cup B) = P(A) + P(B) = \frac{12}{30} + \frac{2}{30} = \frac{14}{30}.$$

(ج) احتمال أنه على الأقل كرة واحده من الكرتين ستكون بيضاء $P(B^c)$

$$P(B^c) = 1 - P(B) = 1 - \frac{2}{30} = \frac{28}{30}.$$

٦-٨-٢ العينات غير المرتبة (Non-Ordered Sample)

تسمى العينة غير مرتبة إذا لم يكن من الضروري الاهتمام بأي من العناصر يظهر قبل الآخر، وبذلك يكون ترتيب ظهور العناصر في العينة المختارة غير مهم.

ويمكن حساب عدد الطرق التي يمكن بها سحب عينة عشوائية غير مرتبة حجمها r من مجموعة من العناصر عددها n من العلاقة التالية:

$$\text{عدد الطرق} = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

مثال (٦-٢٤)

ما هو عدد الطرق التي يمكن بها تكوين بعثة مكونة من 3 رجال وأمرأتين من بين 6 رجال، 5 نساء.

الحل

$$\text{عدد الطرق} = \binom{n_1}{r_1}\binom{n_2}{r_2} = \binom{6}{3}\binom{5}{2} = 20 \times 10 = 200$$

مثال (٦-٢٥)

إذا كان منسوبو قسم الرياضيات بكلية العلوم والدراسات الإنسانية بجامعة الخرج عبارة عن 4 من السعوديين، 7 من المصريين، 2 من الأردنيين، وأردنا تكوين لجنة من ثلاثة أفراد لمتابعة سير الاختبارات الفصلية وتقويمها. فأوجد احتمال

(أ) أن تحتوى اللجنة على سعوديين ومصري.

(ب) أن تحتوى اللجنة على سعودي ومصري وأردني.

(ج) أن لا تحتوى اللجنة على مصريين.

الحل

عدد أفراد القسم 13 عضواً، فإن عدد الطرق اللازمة لتكوين لجنة من ثلاثة أفراد لمتابعة سير الاختبارات الفصلية وتقويمها هي:

$$n(S)=\binom{n}{r}=\binom{13}{3}=286$$

(أ) بفرض أن A حادثة أن اللجنة تحتوى على سعوديين ومصري فإن

$$n(A)=\binom{4}{2}\binom{7}{1}=6\times 7=42,$$

$$P(A)=\frac{n(A)}{n(S)}=\frac{42}{286}.$$

(ب) بفرض أن B حادثة أن تحتوى اللجنة على سعودي ومصري وأردني فإن

$$n(B)=\binom{4}{1}\binom{7}{1}\binom{2}{1}=4\times 7\times 2=56,$$

$$P(B)=\frac{n(B)}{n(S)}=\frac{56}{286}.$$

(ج) بفرض أن C حادثة أن لا تحتوى اللجنة على مصريين فإن

$$n(C)=\binom{4}{3}+\binom{4}{2}\binom{2}{1}+\binom{4}{1}\binom{2}{2}$$

$$=4+6\times 2+4\times 1=20,$$

$$P(C)=\frac{n(C)}{n(S)}=\frac{20}{286}.$$

٦ –٩ الاحتمال الشرطي والاستقلال Conditional probability and Independence

لقد تعرضنا لتعريف الاحتمال وكيفية تعيينه، ولكننا سوف نتعرض هنا لمفهوم الاحتمال الشرطي وتطبيقاته والاستقلال، وهذه المفاهيم تلعب دورا أساسياً في نظرية الاحتمالات. دراسة الاحتمال الشرطي توضح لنا كيف أن إضافة معلومات جديدة تغير من قيمة احتمالات الحوادث. كما أن مفهوم الحوادث المستقلة يرتبط ارتباطاً وثيقاً بالاحتمال الشرطي. ولتقديم مفهوم الاحتمال الشرطي نأخذ المثال التوضيحي التالي:

إذا كان لدينا مجموعة مكونة من أربعة رجال وثلاث نساء، فإذا تم اختيار شخصين من بينهم عشوائياً وعلى التوالي. فما هو احتمال أن يكون الشخص الثاني رجلاً؟

سنرى أن الإجابة على هذا السؤال تعتمد على معلومتنا عن الشخص الأول هل هو رجل أم امرأة؟

- إذا كان الشخص الأول رجلاً فيكون باقياً لدينا في المجموعة ثلاث رجال وثلاث نساء وعلى ذلك يكون احتمال اختيار رجل في المرة الثانية هو 3/6
- أما إذا كان الشخص الأول امرأة فيكون باقياً لدينا في المجموعة أربعة رجال وامرأتان وعلى ذلك يكون احتمال اختيار رجل في المرة الثانية هو 4/6

أي أن كون الشخص الأول رجلاً أو امرأة يؤثر على احتمال اختيار رجل في المرة الثانية، أي أن إضافة معلومات جديدة تؤثر على احتمال الحادثة.

٦-٩-١ الاحتمال الشرطي (Conditional Probability)

إذا كانت A, B حادثتين معرفتين على نفس فضاء العينة S فإن احتمال حدوث الحادثة A بمعلومية حدوث الحادثة B يسمى بالاحتمال الشرطي ويرمز له بالرمز $P(A|B)$ ويعطى من العلاقة:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0$$

مثال (٦-٢٦)

ألقيت زهرتا نرد مرة واحدة فإذا كان مجموع ما سيظهر على حجري النرد 6 فما احتمال أن يكون على إحدى الزهرتين الرقم 2.

الحل

بفرض أن S هو فضاء العينة فإن

$$n(S) = 6^2 = 36$$

وبفرض أن A حادثة أن مجموع النقاط على الحجرين 6،

B حادثة أن أحد الحجرين عليه الرقم 2 فإن

$$A = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$$

$$B = \{(1,2), (2,2), (3,2), (4,2), (5,2), (6,2), (2,1), (2,3), (2,4), (2,5), (2,6)\}$$

$$A \cap B = \{(2,4),(4,2)\}$$

$$\Longrightarrow P(A) = \frac{5}{36},\ \ P(B) = \frac{11}{36}, P(A \cap B) = \frac{2}{36}$$

إذا كان مجموع ما سيظهر على حجري النرد 6 فإن احتمال أن يكون على إحدى الزهرتين الرقم 2 يساوي

$$\therefore P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{2/36}{5/36} = \frac{2}{5}$$

خواص الاحتمال الشرطي:

يحقق الاحتمال الشرطي نفس الخواص التي يحققها الاحتمال المطلق حيث:

١- لأي حادثة A فإن $0 \le P(A|B) \le 1$

٢- $P(\Phi|A) = 0$

٣- $P(S|A) = 1$

٤- لأي حادثتين A, B معرفتين على نفس فضاء العينة وأيضاً حادثة أخرى C بحيث أن $0 < P(C)$ فإن

$$P[(A \cup B)|C] = P(A|C) + P(B|C) - P[(A \cap B)|C]$$

٥- إذا كانت $A_1, A_2, \ldots, A_n$ مجموعة من الحوادث المتنافية المعرفة على نفس فضاء العينة فإن

$$P[(A_1 \cup A_2 \cup \ldots \cup A_n)|B] = \sum_{i=1}^{n} P(A_i|B)$$

٦- $P(A|B) + P(A^C|B) = 1$

قاعدة ضرب الاحتمالات (The multiplication Rule):

لأي حادثتين A, B معرفتين على نفس فضاء العينة لتجربة ما بحيث $0 < P(A) > 0,\ P(B)$ فإنه باستخدام قانون الاحتمال الشرطي يمكن التعبير عن تقاطع الحادثين A, B كحاصل ضرب احتمالين على الصورة

$$P(A \cap B) = P(A)P(B|A)$$
$$= P(B)\,P(A|B)$$

٦-٩-٢ الحوادث المستقلة (Independent Events)

من الواضح أن هناك فرقاً بين احتمال الحادثة A واحتمال A بشرط (بمعلومية) حدوث B أي أن هناك فرقاً بين $P(A|B)$, $P(A)$. ولكن إذا حدث أن $P(A|B) = P(A)$ فإن هذا يعنى أن احتمال وقوع الحادثة A لا يعتمد ولا يتأثر بوقوع أو عدم وقوع الحادثة B وفى هذه الحالة يقال أن A مستقلة عن B ويكون

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

$$\Longrightarrow P(A \cap B) = P(A)\, P(B)$$

وبالتالي فإذا كانت A مستقلة عن B فإن B ستكون مستقلة عن A (وضح).

تعريف:

يقال إن الحادثتين $A,\ B$ حادثتان مستقلتان إذا كان أي من الشروط التالية محققاً:

١- $P(A|B) = P(A)$

٢- $P(B|A) = P(B)$

٣- $P(A \cap B) = P(A)\, P(B)$

نتيجة (٦-٥)

إذا كانت A, B حادثتين مستقلتين فإن

(أ) الحادثتين A, B^C مستقلتان.

(ب) الحادثتين A^C, B مستقلتان.

(ج) الحادثتين A^C, B^C مستقلتان.

الإثبات

بما أن الحادثتين A, B مستقلتان فإن

$$P(A \cap B) = P(A)\, P(B)$$

(أ) الحادثتين A, B^C تكونان مستقلتين إذا تحقق الشرط

$$P(A \cap B^C) = P(A)P(B^C)$$

$$L.H.S. = P(A \cap B^C) = P(A) - P(A \cap B)$$
$$= P(A) - P(A)P(B)$$
$$= P(A)[1 - P(B)]$$
$$= P(A)P(B^C) = R.H.S$$

(ب) الحادثتين A^C, B تكونان مستقلتين إذا تحقق الشرط

$$P(A^C \cap B) = P(A^C)P(B)$$

$$L.H.S. = P(A^C \cap B) = P(B) - P(A \cap B)$$
$$= P(B) - P(A)P(B)$$
$$= P(B)[1 - P(A)]$$
$$= P(B)P(A^C) = R.H.S.$$

(ج) الحادثتين A^C, B^C تكونان مستقلتين إذا تحقق الشرط

$$P(A^C \cap B^C) = P(A^C)P(B^C)$$

$$L.H.S = P(A^C \cap B^C) = P(A \cup B)^C = 1 - P(A \cup B)$$
$$= 1 - [P(A) + P(B) - P(A \cap B)]$$
$$= 1 - P(A) - P(B) + P(A)P(B)$$
$$= P(A^C) - P(B)[1 - P(A)]$$
$$= P(A^C) - P(B)P(A^C)$$
$$= P(A^C)[1 - P(B)]$$
$$= P(A^C)P(B^C) = R.H.S.$$

مثال(٦-٢٧)

من المعلوم أن طالباً سوف ينجح في اختبار القدرات باحتمال 0.9 فإذا قرر ثلاثة زملاء دخول اختبار القدرات. أوجد احتمال أنه سينجح واحد منهم على الأقل في اختبار القدرات.

الحل

بفرض أن B_1 حادثة أن الطالب الأول سوف ينجح في الاختبار،

B_2 حادثة أن الطالب الثاني سوف ينجح في الاختبار،

B_3 حادثة أن الطالب الثالث سوف ينجح في الاختبار،

A حادثة أنه على الأقل واحد من الثلاثة سوف ينجح في الاختبار،

$$P(B_1) = P(B_2) = P(B_3) = 0.9, \qquad P(B_1^C) = P(B_2^C) = P(B_3^C) = 0.1$$

ونلاحظ أن نجاح طالب في الاختبار لا يؤثر على نجاح طالب أخر وبالتالي فإن الحوادث B_1, B_2, B_3 هي حوادث مستقلة وأيضاً

$$A = B_1 \cup B_2 \cup B_3$$

$$P(A) = P(B_1 \cup B_2 \cup B_3)$$

$$= 1 - P(B_1 \cup B_2 \cup B_3)^C$$

$$= 1 - P(B_1^C \cap B_2^C \cap B_3^C)$$

وحيث إن B_1, B_2, B_3 حوادث مستقلة وبالتالي فإن B_1^C, B_2^C, B_3^C ستكون حوادث مستقلة

$$P(A) = 1 - P(B_1^C)P(B_2^C)P(B_3^C)$$

$$= 1 - (0.1)^3 = 0.999$$

مثال(٦–٢٨)

ثلاثة باحثين يقوم كلٌ منهم على حدة بدراسة مشكلة معينة في الصناعة بغرض الوصول لحل لها. وكان احتمال أن يصل الباحث الأول إلى حل هو 2/3 ، واحتمال أن يصل الباحث الثاني إلى حل هو 3/4، واحتمال أن يصل الباحث الثالث إلى حل هو 4/5. احسب احتمال الحوادث التالية:

(أ) التوصل إلى حل.

(ب) عدم التوصل لحل.

(ج) توصل باحث واحد فقط لحل.

(د) توصل باحثين فقط لحل.

الحل

بفرض أن A حادثة توصل الباحث الأول لحل،

B حادثة توصل الباحث الثاني لحل،

C حادثة توصل الباحث الثالث لحل،

فإن

$$P(A) = \frac{2}{3},\ P(B) = \frac{3}{4}, P(C) = \frac{4}{5}$$

(أ) التوصل لحل هو $A \cup B \cup C$

$$\therefore P(A \cup B \cup C) = 1 - P(A \cup B \cup C)^C$$
$$= 1 - P(A^C \cap B^C \cap C^C)$$
$$= 1 - P(A^C)P(B^C)P(C^C)$$
$$= 1 - \frac{1}{3} \times \frac{1}{4} \times \frac{1}{5} = 1 - \frac{1}{60} = \frac{59}{60}$$

(ب) عدم التوصل لحل هو $A^C \cap B^C \cap C^C$

$$\therefore P(A^C \cap B^C \cap C^C) = P(A^C)P(B^C)P(C^C)$$
$$= \frac{1}{3} \times \frac{1}{4} \times \frac{1}{5} = \frac{1}{60}$$

(ج) توصل باحث فقط لحل $[(A \cap B^C \cap C^C) \cup (A^C \cap B \cap C^C) \cup (A^C \cap B^C \cap C)]$

$$\therefore P[(A \cap B^C \cap C^C) \cup (A^C \cap B \cap C^C) \cup (A^C \cap B^C \cap C)] =$$
$$P(A \cap B^C \cap C^C) + P(A^C \cap B \cap C^C) + P(A^C \cap B^C \cap C) =$$
$$P(A)P(B^C)P(C^C) + P(A^C)P(B)P(C^C) + P(A^C)P(B^C)P(C) =$$
$$= \frac{2}{3} \times \frac{1}{4} \times \frac{1}{5} + \frac{1}{3} \times \frac{3}{4} \times \frac{1}{5} + \frac{1}{3} \times \frac{1}{4} \times \frac{4}{5} = \frac{9}{60}$$

(د) توصل باحثين فقط لحل $[(A \cap B \cap C^C) \cup (A \cap B^C \cap C) \cup (A^C \cap B \cap C)]$

$$\therefore P[(A \cap B \cap C^C) \cup (A \cap B^C \cap C) \cup (A^C \cap B \cap C)] =$$
$$P(A \cap B \cap C^C) + P(A \cap B^C \cap C) + P(A^C \cap B \cap C) =$$
$$P(A)P(B)P(C^C) + P(A)P(B^C)P(C) + P(A^C)P(B)P(C) =$$

$$= \frac{2}{3}\times\frac{3}{4}\times\frac{1}{5}+\frac{2}{3}\times\frac{1}{4}\times\frac{4}{5}+\frac{1}{3}\times\frac{3}{4}\times\frac{4}{5}=\frac{26}{60}$$

مثال(٦-٢٩)

يقوم أحد الموظفين بالمرور في طريقه إلى عمله بثلاث إشارات ضوئية تعمل بصورة مستقلة فإذا كان احتمال أن تكون الإشارات الثلاث خضراء هو 0.7, 0.6, 0.8 على الترتيب. ما هو احتمال أن يمر هذا الشخص بالإشارات الثلاث دون توقف؟

الحل

بفرض أن

A: حادثة أن الإشارة الأولى خضراء،

B: حادثة أن الإشارة الثانية خضراء،

C: حادثة أن الإشارة الثالثة خضراء،

D: حادثة أن يمر الشخص بدون توقف في الإشارات الثلاث.

فإن

$$P(A)=0.7,\qquad P(B)=0.6,\quad P(C)=0.8$$

الشخص سيمر بدون توقف في الإشارات الثلاث إذا كانت الإشارات الثلاث خضراء لحظة وصوله إليها.

$$P(D)=P(A\cap B\cap C)=P(A)P(B)P(C)$$
$$=0.7\times0.6\times0.8=0.336$$

٦-٩-٣ قانون الاحتمال الكلي (Law of Total Probability)

إذا كانت $H_1, H_2, \ldots, H_n$ مجموعة من الحوادث المتنافية مثنى مثنى والمعرفة على نفس فضاء العينة لتجربة عشوائية ما بحيث

$$\cup_{i=1}^{n} H_i = S,\qquad H_i\cap H_j=\Phi,\qquad \forall\, i\neq j$$

وكانت الحادثة A معرفة على نفس فضاء العينة. فإن

$$P(A)=\sum_{i=1}^{n} P(A|H_i)P(H_i)$$

الإثبات

حيث إن الحوادث $H_1, H_2, \ldots, H_n$ مجموعة من الحوادث المتنافية ويمثل اتحادها كل فضاء العينة. فإن أي حادثة A يمكن كتابتها على الصورة

$$A = [(A \cap H_1) \cup (A \cap H_2) \cup \ldots \cup (A \cap H_n)]$$

حيث أن كل الحوادث $H_1, H_2, \ldots, H_n$ حوادث متنافية فإن الحوادث $A \cap H_1, A \cap H_2, \ldots, A \cap H_n$ أيضاً حوادث متنافية

$$P(A) = P[(A \cap H_1) \cup (A \cap H_2) \cup \ldots \cup (A \cap H_n)]$$
$$= P(A \cap H_1) + P(A \cap H_2) + \ldots + P(A \cap H_n)$$

ومن قاعدة ضرب الاحتمالات نجد أن

$$P(A \cap H_i) = P(A|H_i)P(H_i)$$

$$\Rightarrow P(A) = P(A|H_1)P(H_1) + P(A|H_2)P(H_2) + \ldots + P(A|H_n)P(H_n)$$
$$= \sum_{i=1}^{n} P(A|H_i)P(H_i)$$

نتيجة (٦-٦)

إذا كانت A حادثة في فضاء العينة S فإنه لكل حادثة أخرى B يكون

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$$

وذلك لأن A^C, A حوادث متنافية واتحادهما يمثل S.

٦-٩-٤ نظرية بيز (Bayes` Theorem)

إذا كانت $H_1, H_2, \ldots, H_n$ مجموعة من الحوادث المتنافية مثنى مثنى والمعرفة على نفس فضاء العينة لتجربة عشوائية بحيث

$$\cup_{i=1}^{n} H_i = S, \qquad H_i \cap H_j = \Phi, \qquad \forall\, i \neq j$$

وكانت الحادثة A معرفة على نفس فضاء العينة. فإن

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_{j=1}^{n} P(A|H_j)P(H_j)}$$

الإثبات

من قانون الاحتمال الشرطي فإن

$$P(H_i|A) = \frac{P(A \cap H_i)}{P(A)}$$

لكن من قانون الاحتمال الكلي فإن

$$P(A) = \sum_{j=1}^{n} P(A|H_j)P(H_j)$$

ومن قاعدة ضرب الاحتمالات

$$P(A \cap H_i) = P(A|H_i)P(H_i)$$

$$\Rightarrow P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_{j=1}^{n} P(A|H_j)P(H_j)}$$

مثال(٦–٣٠)

ثلاثة ماكينات I, II, III تنتج على الترتيب 20% ,30% ,50% من الإنتاج الكلي لأحد المصانع. وكانت نسبة إنتاج قطعة تالفة في الماكينات الثلاثة I, II, III هي 5% ,4% ,3% على الترتيب. سحبت قطعة من الإنتاج الكلى عشوائياً.

(أ) أوجد احتمال أن تكون القطعة المسحوبة تالفة.

(ب) إذا كانت القطعة المسحوبة تالفة، ما احتمال أن تكون من إنتاج الماكينة III؟

الحل

بفرض أن

A حادثة أن القطعة المسحوبة تالفة،

H_1 القطعة المسحوبة من إنتاج الماكينة I،

H_2 القطعة المسحوبة من إنتاج الماكينة II،

H_3 القطعة المسحوبة من إنتاج الماكينة III.

فإن

$$P(H_1) = 0.5, \quad P(H_2) = 0.3, \quad P(H_3) = 0.2$$

$$P(A|H_1) = 0.03, \quad P(A|H_2) = 0.04, \quad P(A|H_3) = 0.05$$

(أ) احتمال أن تكون القطعة تالفة

من قانون الاحتمال الكلي فإن

$$P(A) = \sum_{j=1}^{3} P(A|H_j)P(H_j)$$

$$= 0.5 \times 0.03 + 0.3 \times 0.04 + 0.2 \times 0.05$$

$$= 0.037$$

(ب) إذا كانت القطعة المسحوبة تالفة، فإن احتمال أن تكون من إنتاج الماكينة III

$$P(H_3|A) = \frac{P(A|H_3)P(H_3)}{P(A)}$$

$$= \frac{0.2 \times 0.05}{0.037} = 0.27$$

مثال(٦-٣١)

قطع تصنع في مصنع معين وتوجه لاختبار صلاحيتها بواسطة اثنين من العمال. فإذا كان احتمال أن توجه القطع إلى العامل الأول هو 60% والعامل الثاني هو 40% واحتمال أن تكون القطعة صالحة بشرط اختبارها بواسطة العامل الأول هو 94% وبواسطة العامل الثاني هو 98% فإذا كانت القطعة في نهاية الاختبار صالحة. أوجد احتمال أن تكون هذه القطعة اختبرت بواسطة العامل الأول.

الحل

بفرض أن A حادثة أن القطعة صالحه،

B_1 حادثة أن القطعة تم اختبارها بواسطة العامل الأول،

B_2 حادثة أن القطعة تم اختبارها بواسطة العامل الثاني

فإن

$$P(B_1) = \frac{60}{100} = 0.6\,, \quad P(B_2) = \frac{40}{100} = 0.4$$

$$P(A|B_1) = \frac{94}{100} = 0.94, \quad P(A|B_2) = \frac{98}{100} = 0.98$$

وباستخدام قاعدة بيز

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{\sum_{j=1}^{2} P(A|B_j)P(B_j)}$$
$$= \frac{0.94 \times 0.6}{0.94 \times 0.6 + 0.98 \times 0.4} = 0.59$$

مثال(٦-٣٢)

ثلاثة فصول في روضة للأطفال، في كلٍّ منها 30 طفلاً. أطفال الفصل الأول كلهم أولاد، وأطفال الفصل الثاني كلهن بنات، أما أطفال الفصل الثالث فنصفهم أولاد. اختير فصل بطريقة عشوائية ومنه اختير طفل.

(أ) فما احتمال أن يكون الطفل المختار ولداً،

(ب) إذا علم أن الطفل المختار ولد، فما احتمال أن يكون من الفصل الأول؟

الحل

بفرض أن A حادثة أن الطفل المختار ولد،

B_1 الطفل المختار من الفصل الأول،

B_2 الطفل المختار من الفصل الثاني،

B_3 الطفل المختار من الفصل الثالث.

فإن

$$P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$$
$$P(A|B_1) = 1, \qquad P(A|B_2) = 0, \qquad P(A|B_3) = \frac{1}{2}$$

(أ) احتمال أن يكون الطفل المختار ولداً

فمن قانون الاحتمال الكلي فإن

$$P(A) = \sum_{j=1}^{3} P(A|B_j)P(B_j)$$

$$= \frac{1}{3} \times 1 + \frac{1}{3} \times 0 + \frac{1}{3} \times \frac{1}{2} = \frac{1}{2}$$

(ب) إذا علم أن الطفل المختار ولد، فما احتمال أن يكون من الفصل الأول

باستخدام قاعدة بيز ينتج أن

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A)} = \frac{\left(\frac{1}{3}\right) \times 1}{\left(\frac{1}{2}\right)} = \frac{2}{3}$$

أسئلة وتمارين (٦)

١- عرف فضاء العينة، الحادثة، الحوادث المتنافية.

٢- اكتب فضاء العينة لعدد مرات رمي قطعة نقود معدنية حتى الحصول على صورة.

٣- إذا كانت $S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

$$A = \{2, 4, 6, 8, 10\}, \quad B = \{1, 3, 5, 7, 9\}, C = \{1, 3, 5\}, D = \{2, 3, 4, 5\}$$

اكتب عناصر الحوادث التالية

$$A \cup B, \quad A \cap B, \quad C^C, \quad (C^C \cap D) \cup B, \quad A \cap C \cap D^C$$

٤- إذا كان احتمال نجاح طالب في مقرر 1080إحص هو 0.79 فما هو احتمال رسوبه؟

٥- إذا كانت A, B حادثتين معرفتين على فضاء العينة S فعبر عن الحوادث التالية باستخدام شكل فن

$$A^C, \; A \cap B, A \cup B, \; A \cap B^C, \; A^C \cap B, B - A, (B - A)^C$$

٦- إذا كانت A, B, C ثلاثة حوادث معرفة على نفس فضاء العينة فعبر عن التعبيرات التالية باستخدام A, B, C

- وقوع A فقط،
- وقوع A, B وعدم وقوع C،
- وقوع حادثة واحدة على الأقل،
- وقوع حادثتين على الأقل،
- وقوع الثلاث حوادث جميعاً،

- عدم وقوع أي من الحوادث الثلاثة،
- على الأكثر حادثة واحدة فقط ستحدث،
- وقوع حادثين فقط.

٧- إذا كان لدينا صندوق يحتوى على كرتين باللون الأبيض فكم عدد الكرات التي يمكن اختيارها من الصندوق إذا كان السحب بإرجاع أو بدون إرجاع؟

٨- بكم طريقة يمكن تكوين لجنة مكونة من 3 أفراد من مجموعة بها 20 شخصاً؟

٩- إذا كان لدينا مجموعة مكونة من 5 سيدات، 7 رجال فبكم طريقة يمكن تكوين لجنة مكونة من سيدتين وثلاث رجال في الحالات التالية:

(أ) إذا كان هناك رجل معين يجب أن يتواجد في اللجنة.

(ب) إذا كان هناك رجلين لا يمكن تواجدهما باللجنة.

(ج) إذا كان كل من الرجال والسيدات في المجموعة لهم حق التواجد باللجنة.

١٠- إذا كان لدينا مجموعة من 10 أطفال بكم طريقة يمكن تقسيمها إلى فريقين بحيث يحتوى كل فريق على 5 أطفال؟

١١- إذا كان لدينا 12 مكعباً ملوناً بحيث يوجد منها 6 سوداء، 4 حمراء، ومكعب واحد ابيض، مكعب واحد ازرق. بكم طريقة يمكن وضع هذه المكعبات في صف؟

١٢- بكم طريقة يمكن ترتيب 3 كتب فيزياء، كتابين رياضيات، كتاب كيمياء على رف للكتب إذا كان

(أ) الكتب توضع بأي ترتيب.

(ب) يجب أن توضع كتب الرياضيات بجوار بعضها وكتب فيزياء بجوار بعضها.

(ج) كتب فيزياء يجب أن توضع بجوار بعضها وباقي الكتب توضع بأي ترتيب.

١٣- إذا تم إلقاء حجري نرد معاً، فما هو احتمال أن يكون مجموع ما سيظهر على الحجرين يساوى 7؟

١٤- إذا أردنا سحب 3 كرات عشوائياً من وعاء يحتوى على 6 كرات بيضاء، 5 كرات سوداء فما هو احتمال أن تكون العينة المسحوبة تحتوى على كرة واحدة بيضاء وكرتين سوداء؟

١٥- أراد المسئولون في المستشفى الجامعي بجامعة الملك سعود تنظيم قواعد البيانات للمرضى فاقترحوا عمل ملف لكل مريض بحيث يمثل كل ملف بكود مكون من 13 خانة بحيث يحتوى الكود على تاريخ إنشاء

الملف مقسم من اليسار إلى اليمين إلى خانتين تعبر عن السنة وخانتين تعبر عن الشهر وخانتين تعبر عن اليوم والباقي يعبر عن رقم المريض مع ملاحظه أنه لا يوجد ملف له الرقم صفر فما هي عدد الملفات التي يمكن أن يتم إنشاؤها في خلال خمسة أعوام قادمة.

١٦- إذا كانت A, B حادثتين متنافيتين وكان $P(A) = 0.4, \quad P(B)= 0.5$ أوجد

$P(A \cup B), P(A^C), P(A \cap B^C), P(B \cap A^C)$

١٧- إذا كان احتمال أن ينجح طالب في الرياضيات هو 2/3 واحتمال أن ينجح في الأحياء هو 4/9 واحتمال أن ينجح في مادة منهما هو 4/5. ما هو احتمال أن ينجح الطالب في كلٍّ من المادتين؟

١٨- ألقيت زهرتا نرد أحدهما حمراء والأخرى خضراء. أكتب فضاء العينة لهذه التجربة ثم أوجد احتمال أن يكون مجموع الرقمين الناتجين زوجياً.

١٩- لأي حادثتين A , B أثبت أن

$$P[(A \cap B^C) \cup (A^C \cap B)] = P(A) + P(B) - 2P(A \cap B)$$

٢٠- ألقيت قطعة عملة 6 مرات متتالية. ما احتمال أن تظهر صورة واحدة على الأقل؟

٢١- ثلاثة فرق A, B, C في مسابقة كأس العالم. فإذا كانت فرصة الفرقة B في الفوز بالكأس ضعف فرصة فوز الفرقة A. كما أن فرصة C في الفوز بالكأس ضعف فرصة فوز B . أوجد احتمال فوز الفريق B أو C بالكأس.

٢٢- تم اختيار 3 كتب عشوائياً من أحد الأرفف في مكتبة وكان يحتوى على 4 كتب رياضيات، 3 كتب إحصاء، وقاموس. أوجد احتمال

(أ) وجود القاموس بين المجموعة التي تم اختيارها.

(ب) الحصول على كتابين رياضيات وكتاب إحصاء.

(ج) عدم وجود القاموس بين المجموعة المختارة.

٢٣- من صندوق يحتوى 3 كرات حمراء، كرتان بيضاء، 4 كرات زرقاء، 5 خضراء. سحبت كرة عشوائياً. أوجد احتمال أن تكون الكرة المسحوبة:

(أ) حمراء أو بيضاء.

(ب) ليست (حمراء أو زرقاء).

٢٤- أرادت وحدة المرور بالرياض تحديد عدد لوحات السيارات التي تحتوى على ثلاثة أحَرُف وثلاثة أرقام مع ملاحظة انه لا توجد سيارة تحمل الرقم صفر. فما هو هذا العدد؟ وإذا قامت بزيادة عدد الحروف إلى أربعة وأيضا الأرقام، فما هو العدد المتوقع من السيارات التي يمكن أن تحمل مثل تلك اللوحات؟

٢٥- صندوق يحتوى 4 كرات بيضاء، 2 حمراء، 3 خضراء، سحبت منه كرتان بإرجاع. أوجد احتمال أن تكون الكرتان المسحوبتان

(أ) بيضاء.

(ب) حمراء.

(ج) واحدة حمراء والأخرى بيضاء.

(د) لهما نفس اللون.

٢٦- تم الاحتياج إلى الدم في عملية لنقله حيث احتوت قائمة من 50 فرداً من المتطوعين بدمائهم على 10 أفراد دمهم من فصيلة B، 15 فرداً دمهم من النوع A، 20 فرداً دمهم من النوع O ، 5 أفراد دمهم من النوع AB إذا تم اختيار أربعة أفراد عشوائياً من القائمة. أوجد احتمال أن يكون

(أ) جميعهم من الفصيلة B

(ب) اثنان من الفصيلة B وواحد من الفصيلة AB

(ج) على الأقل واحد من الفصيلة A

(د) اثنان على الأكثر من الفصيلة B

(هـ) جميع الفصائل ممثله في العينة المختارة

٢٧- في تجربة إلقاء حجري نرد معاً فما هو احتمال أن يظهر على الحجر الثاني رقم اكبر مما سيظهر على الحجر الأول؟

٢٨- إذا علمت أن الرقم 3 لم يظهر عند إلقاء زهرة النرد مرة واحدة. فما هو احتمال أن يكون الرقم الذي حصلنا عليه رقماً فردياً؟.

٢٩- ثلاثة من لاعبي كرة القدم يصوبون بالكرة نحو المرمى من نقطة ضربة الجزاء. فإذا كان احتمال أن يحرز كل منهم هدفاً هو 1/4 ,1/3 ,1/2 على الترتيب. أوجد احتمال أن يحرز واحد منهم فقط هدفاً. وإذا أحرز

واحد منهم فقط الهدف فما هو احتمال أن يكون هو اللاعب الأول (علماً بأن لكل لاعب الحق في تصويبة واحدة فقط)؟.

٣٠- إذا كانت الحادثتان A, B حادثتين معرفتين على فضاء العينة S بحيث

$$P(A) = \frac{1}{4}, \quad P(A \cup B) = \frac{1}{3}, \quad P(B) = K$$

أوجد قيمة K في الحالات التالية:

(أ) A, B حادثتين متنافيتين.

(ب) A, B حادثتين مستقلتين

٣١- صندوق I يحتوى 3 كرات حمراء، 6 كرات بيضاء وصندوق II يحتوى 5 كرات حمراء، 2 بيضاء. ألقيت زهرة نرد بحيث إذا كانت النتيجة 1 أو 3 تم سحب كرة من الصندوق I وبخلاف ذلك تسحب الكرة من الصندوق II

(أ) أوجد احتمال أن تكون الكرة المسحوبة حمراء؟

(ب)إذا علم أن الكرة المسحوبة حمراء. ما هو احتمال أن تكون قد أخذت من الصندوقI؟

(ج) إذا علم أن الكرة المسحوبة بيضاء. ما هو احتمال أن تكون قد أخذت من الصندوقI؟

٣٢- إذا كانت $P(A) = \frac{1}{3}, P(B) = \frac{1}{2}, P(A \cup B) = \frac{3}{4}$ أوجد قيم الاحتمالات $P(A|B), P(A|B^C)$, $P(A^C|B), P(A^C|B^C)$

٣٣- صندوق يحتوى 7 كرات بيضاء مرقمة من 1 إلى 7 وأيضا 7 كرات حمراء مرقمة من 1 إلى 7 سحبت منه كرة واحدة عشوائياً. إذا علمت أن الكرة المسحوبة عليها رقم زوجي فما هو احتمال أن تكون حمراء؟

٣٤- لنفرض أن عدد الذكور يساوى عدد الإناث في إحدى المدن وأن 7% من الإناث، 30% من الذكور مصابون بمرض ارتفاع ضغط الدم، فإذا اخترنا أحد السكان عشوائياً

(أ) ما هو احتمال أن يكون مصاباً.

(ب) إذا كان مصاباً، فما هو احتمال أن يكون أنثى.

(ج) أوجد الاحتمالات السابقة بفرض أن عدد الذكور ضعف عدد الإناث.

٣٥- احتمال أن يصيب قناص الهدف في أي رمية هو 0.85 واحتمال إصابته للهدف في أي رمية مستقلة عن الرميات الأخرى، إذا كان لديه مسدسان أحدهما يحتوى رصاصة واحدة فقط، والأخر يحتوى على

رصاصتين، فإذا اختار أحد المسدسين بشكل عشوائي وأطلق على الهدف حتى أفرغ الرصاص الموجود فيه، احسب احتمال أن يصيب الهدف مرة واحدة فقط.

٣٦- ممر به أربع بوابات تعمل بشكل مستقل فإذا كان احتمال أن يجد شخص يمر بهذا الممر أي بوابه مفتوحة هو 0.5 فما هو احتمال أن يجد جميع البوابات مفتوحة؟

٣٧- إذا كان $P(C)=\frac{1}{3}, P(D|C)=\frac{1}{2}, P(C\cup D)=\frac{4}{5}$ فهل C, D مستقلتان؟

٣٨- إذا كانت A, B مستقلتين، وكان $P(A\cup B)=0.75$ ، $P(A)=0.25$ فأوجد P(B) ثم احسب $P(A|B^c), P(A^c|B^c), P(A^c|B)$

٣٩- إذا كانت A, B حادثتين متنافيتين وكانت P(A)=0.3 ، P(B)=0.45 فاحسب الاحتمالات التالية: $P(A|B), P(A\cup B), P(A^c\cup B^c), P(A^c\cap B)$

٤٠- إذا كانت $P(A\cap B^c)=0.2$ ، $P(A^c)=0.7$ فاحسب احتمال $P(A|B)$

٤١- إذا كانت A, B حادثتين مستقلتين وكانت P(A) =0.4 ، $P(A\cup B)=0.6$ فاحسب $P(B|A)$

٤٢- إذا كان P(A)=0.3، $P(A\cup B)=0.8$ فأوجد P(B|A) في كل من الحالات التالية:

(أ) A, B حادثتان مستقلتان.

(ب) A, B حادثتان متنافيتان.

(ج) الحادثة B حادثة جزئية من الحادثة A

٤٣- من المعلوم أن مريضاً سوف يستجيب لعلاج مرض معين باحتمال 0.9 فإذا تلقى ثلاثة مرضى هذا العلاج بطريقة مستقلة. أوجد احتمال أن يستجيب واحد منهم على الأقل لهذا العلاج.

٤٤- في تجربة إلقاء حجري نرد معاً ما هو احتمال أن يظهر الرقم 6 على الأقل على حجر نرد واحد إذا علمنا أن الرقمين الظاهرين على الحجرين مختلفان؟

الفصل السابع

المتغيرات العشوائية

RANDOM VARIABLES

٧-١ مقدمة

لقد تعرضنا في الفصل السابق لبعض مفاهيم الاحتمالات والتجارب العشوائية والحوادث. لكن في كثير من الأحيان نرغب في التعامل مع قيم عددية مرتبطة بنقاط العينة العشوائية بدلاً من التعامل مع نقاط العينة نفسها. لذا فإننا سنقوم بتحويل هذه القيم الوصفية إلى قيم عددية حقيقية تسمى قيم المتغير العشوائي. فإن المتغيرات العشوائية تستخدم للتعبير عن نتائج التجربة العشوائية وعن الحوادث بقيم عددية بدلاً من مسميات أو صفات. فمثلاً في تجربة إلقاء قطعتي عملة معاً مرة واحدة فإن فراغ العينة هو:

$$S = \{(H,H),(H,T),(T,H),(T,T)\}$$

وبفرض أن X ترمز لعدد الصور المحتملة كنتائج للتجربة. وعليه فإن X متغير يمكن أن يأخذ القيم 0 أو 1 أو 2 وهو بالتالي متغير عشوائي نظراً لعدم إمكانية تحديد قيمته مقدماً أي قبل أن يتم الانتهاء من التجربة. قيمة $X=1$ تعنى ظهور الصورة مرة واحدة ويمثلها الحدث المركب $\{(H,T), (T,H)\}$. وقيمة $X=2$ تمثل الحدث البسيط $\{(H,H)\}$. وقيمة $X=0$ تمثل الحدث البسيط $\{(T,T)\}$. وأن $X\leq 2$ تمثل فراغ العينة كله.

٧-٢ المتغير العشوائي (Random Variable)

تعريف (٧-١)

المتغير العشوائي هو متغير يأخذ قيماً تعتمد على الصدفة بناء على فراغ العينة لتجربة عشوائية.

بناء على التعريف السابق يمكننا معرفة بعض الأمثلة في حياتنا اليومية للمتغيرات العشوائية مثل:

- وقت الانتظار الذي يقضيه شخص ما لإنهاء معاملاته داخل أحد مراكز الخدمة (بنك – سوق مركزيه – ورشة لإصلاح السيارات – مكتب بريد.......)
- عدد الطلبة الحاضرين في أحد الفصول الدراسية.
- عدد أفراد الأسرة في مدينة ما.
- درجات مجموعة من الطلاب في اختبار ما.
- عدد القطع التالفة في عينة عشوائية حجمها 10 (مثلاً) مسحوبة من الإنتاج الكلى لأحد المصانع.
- أطوال مجموعة من الطلاب.
- مستوي الهيموجلوبين لمرضى السكر

مثال (٧-١)

ألقيت قطعة عملة ثلاث مرات متتالية. وكان المتغير العشوائي X يمثل عدد مرات ظهور صورة. وضح القيم الممكنة للمتغير العشوائي.

الحل

فضاء العينة لهذه التجربة هو

$$S = \{(H,H,H),(H,H,T),(H,T,H),(H,T,T),(T,H,H),(T,H,T),(T,T,H),(T,T,T)\}$$

ويمكن تعيين قيم المتغير العشوائي X كما يلي:

عدد الصور	نقطة العينة
3	(H,H,H)
2	(H,H,T)
2	(H,T,H)
1	(H,T,T)
2	(T,H,H)
1	(T,H,T)
1	(T,T,H)
0	(T,T,T)

فإن القيم الممكنة للمتغير العشوائي X والتي تمثل عدد مرات ظهور صورة هي $0, 1, 2, 3$

ويمكن تعريف المتغير العشوائي رياضياً بالصورة التالية.

تعريف (٧-٢)

المتغير العشوائي هو دالة حقيقية معرفة على فضاء العينة لتجربة عشوائية.

ونلاحظ في المثال السابق أنه تم تعريف المتغير العشوائي X على فضاء العينة لتجربة إلقاء قطعة العملة ثلاثة مرات وكانت قيم المتغير العشوائي هي قيماً حقيقية وهى 3 ,2 ,1 ,0 لذا يمكن تعريف المتغير X على الصورة التالية:

$$X: S \longrightarrow \mathcal{R}$$

ونرمز دائما للمتغير العشوائي بالأحرف الكبيرة ... ,X, Y, Z وقيم المتغير العشوائي بالأحرف الصغيرة ... ,x, y, z

٧-٣ أنواع المتغيرات العشوائية

المتغير العشوائي يتحدد نوعه بناء على القيم التي يمكن أن يأخذها ذلك المتغير. ويوجد نوعان من المتغيرات العشوائية هي:

(أ) المتغير العشوائي المتقطع (Discrete Random Variable)

هو المتغير الذي يمكن أن يأخذ عدداً محدوداً أو عدداً لا نهائياً من القيم المنفصلة.

ومن أمثلة المتغير العشوائي المنفصل:

- عدد الأبناء الذكور لدي السيدة السعودية،
- عدد مرات إلقاء قطعة العملة حتى الحصول على صورة لأول مرة،
- عدد المكالمات التي تصل إلى سنترال معين خلال اليوم،
- عدد الأشخاص الذي يستعملون صرافاً آلياً معيناً في اليوم،
- عدد الطلاب الحاضرين محاضرات الإحصاء في يوم ما.
- وغير ذلك.

(ب) المتغير العشوائي المتصل (Continuous Random Variable)

هو المتغير الذي يمكن أن يأخذ أية قيمة داخل فترة محدودة أو غير محدودة.

ومن أمثلة المتغير العشوائي المتصل:

- عمر الفرد السعودي،
- الزمن الذي يستغرقه موظف للوصول من محل إقامته لمقر عمله،
- الزمن الذي تستغرقه مكالمة تليفونية تصل لسنترال ما خلال اليوم،
- وقت إصلاح أجهزة الحاسب الموجودة في الكلية.
- وغير ذلك.

٧-٤ وصف المتغير العشوائي (Random Variable Description)

يتحدد المتغير العشوائي تماماً متى عرفت القيم التي يمكن أن يأخذها بالإضافة إلى معرفة الاحتمالات المناظرة لها. وهذه الاحتمالات المناظرة لقيم المتغير العشوائي تسمى بالوصف الاحتمالي للمتغير.

٧-٤-١ دالة التوزيع للمتغير العشوائي (Distribution Function)

كل متغير عشوائي X يناظره دالة توزيع $F(x)$ والتي تعرف بالصورة التالية:

$$F(x) = P[X \leq x], \quad x \in X(S)$$

خواص دالة التوزيع:

١- توجد دالة التوزيع للمتغيرات العشوائية المتصلة والمتقطعة.

٢- دالة التوزيع دالة غير تناقصية.

٣- دالة التوزيع تحقق العلاقات التالية:

(i) $0 \leq F(x) \leq 1$,

(ii) $F(\infty) = \lim_{x\to\infty} F(x) = 1$

(iii) $F(-\infty) = \lim_{x\to-\infty} F(x) = 0$

(iv) $P[a < X \leq b] = F(b) - F(a), \quad a < b$

وسوف نستعرض المتغيرات العشوائية المنفصلة والمتصلة ببعض من التفصيل.

٧-٤-٢ المتغير العشوائي المتقطع(المنفصل) (Discrete Random Variable)

لقد عرفنا من قبل المتغير العشوائي المتقطع لكننا هنا سوف نهتم بوصف هذا النوع من المتغيرات العشوائية.

التوزيع الاحتمالي (Probability Distribution):

بفرض أن المتغير العشوائي X يمكن أن يأخذ القيم المنفصلة $x_1, x_2, \ldots, x_n, \ldots$ فيعرف التوزيع الاحتمالي والذي يطلق عليه دالة الكتلة الاحتمالية للمتغير العشوائي X بأنه مجموعة الاحتمالات المناظرة لقيم المتغير العشوائي وتعرف على الصورة

$$p(x) = P[X = x],\ \ x = x_1, x_2, x_3, \ldots$$

والذي يحقق الخواص التالية:

(i) $0 \leq p(x) \leq 1$

(ii) $\sum_{x} p(x) = 1$

(iii) $F(x) = P[X \leq x] = \sum_{t \leq x} p(t)$

ويمكن كتابة دالة الكتلة الاحتمالية عندما يكون عدد القيم الممكنة للمتغير العشوائي X محدوداً وصغيراً على صورة جدول كالتالي

X	x_1	x_2	...	x_n	المجموع
$p(x)$	$p(x_1)$	$p(x_2)$	...	$p(x_n)$	1

مثال (٧-٢)

صندوق يحتوى 4 كرات بيضاء، 3 كرات سوداء. سحبت منه بدون إرجاع ثلاث كرات. وكان المتغير X يمثل عدد الكرات البيضاء في العينة.

(أ) أوجد التوزيع الاحتمالي للمتغير X ومثله بيانياً.

(ب) أوجد احتمال أن تحتوى العينة على كرة بيضاء واحده على الأقل.

(ج) أوجد دالة التوزيع $F(x)$ ومثلها بيانياً.

الحل

بفرض أن X يمثل عدد الكرات البيضاء في العينة، فنلاحظ أن القيم الممكنة للمتغير X هي $0, 1, 2, 3$

3	4
سوداء	بيضاء

(أ) ولتعيين التوزيع الاحتمالي يجب معرفة الاحتمالات المناظرة لقيم المتغير X كالتالي:

$$p(0) = P[X = 0] = P\left[\text{الكرات المختارة سوداء}\right] = \frac{3(2)(1)}{7(6)(5)} = 0.029$$

$$p(1) = P[X = 1] = P\left[\text{كرة واحدة بيضاء}\right] = \frac{4(3)(2) + 3(4)(2) + 3(2)(4)}{7(6)(5)} = 0.343$$

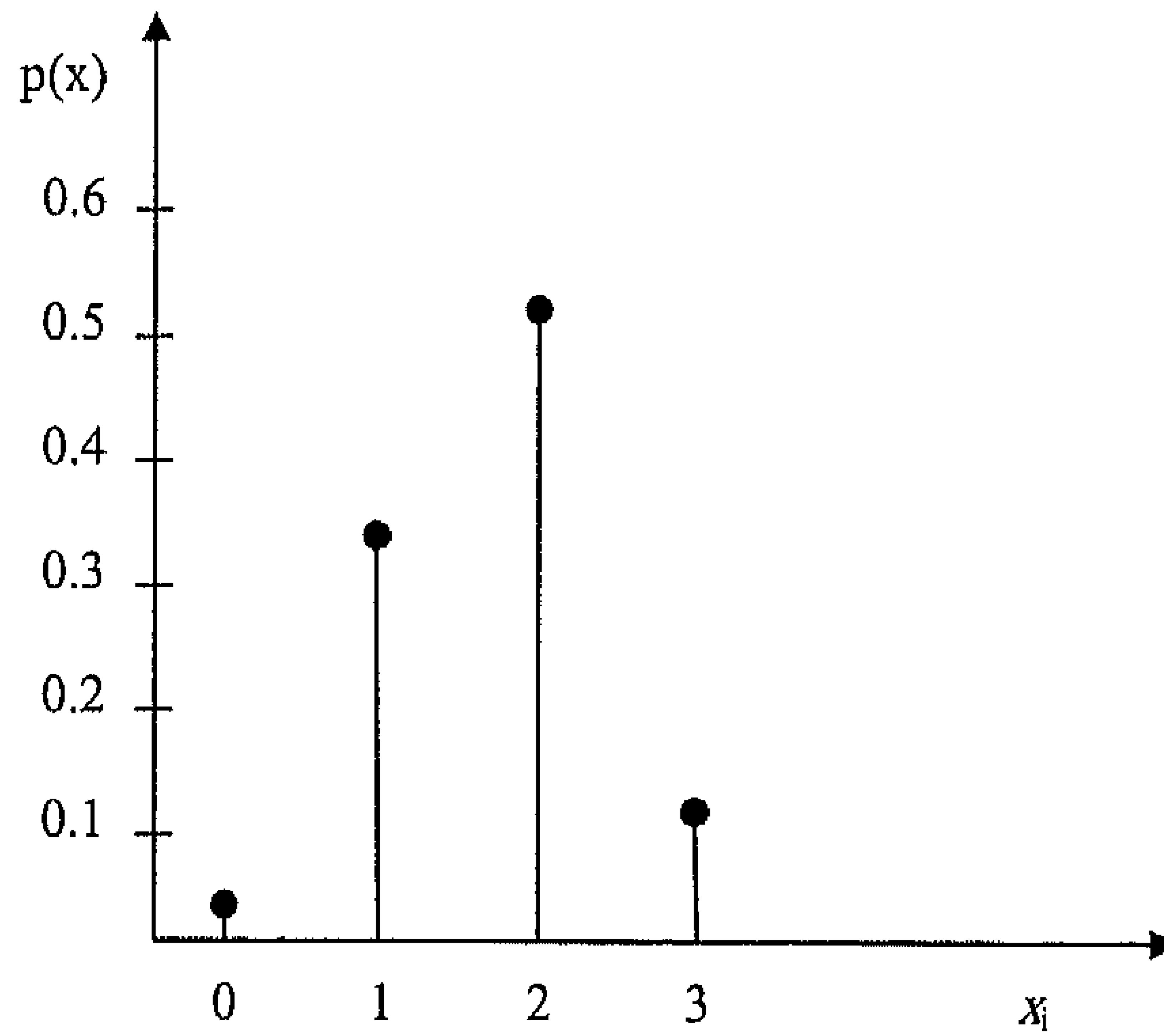
$$p(2) = P[X = 2] = P\left[\text{كرتين بيضاء}\right] = \frac{4(3)(3) + 4(3)(3) + 3(4)(3)}{7(6)(5)} = 0.514$$

$$p(3) = P[X = 3] = P\left[\text{الكرات جميعها بيضاء}\right] = \frac{4(3)(2)}{7(6)(5)} = 0.114$$

ويصبح التوزيع الاحتمالي (دالة الكتلة الاحتمالية) للمتغير X على الصورة

X	0	1	2	3	المجموع
$p(x)$	0.029	0.343	0.514	0.114	1

ويمكن تمثيل التوزيع الاحتمالي بالصورة البيانية التالية:



(ب) احتمال أنه على الأقل كرة واحدة بيضاء

$$P[X \geq 1] = 1 - P[X < 1]$$
$$= 1 - P[X = 0] = 1 - p(0)$$
$$= 1 - 0.029 = 0.971$$

(ج) يمكن حساب دالة التوزيع كالتالي:

$$F(x) = P[X \leq x], \ \ x = 0, 1, 2, 3.$$

$$F(0) = P[X \leq 0] = P[X = 0] = p(0) = 0.029$$

$$F(1) = P[X \leq 1] = p(0) + p(1) = 0.029 + 0.343 = 0.372$$

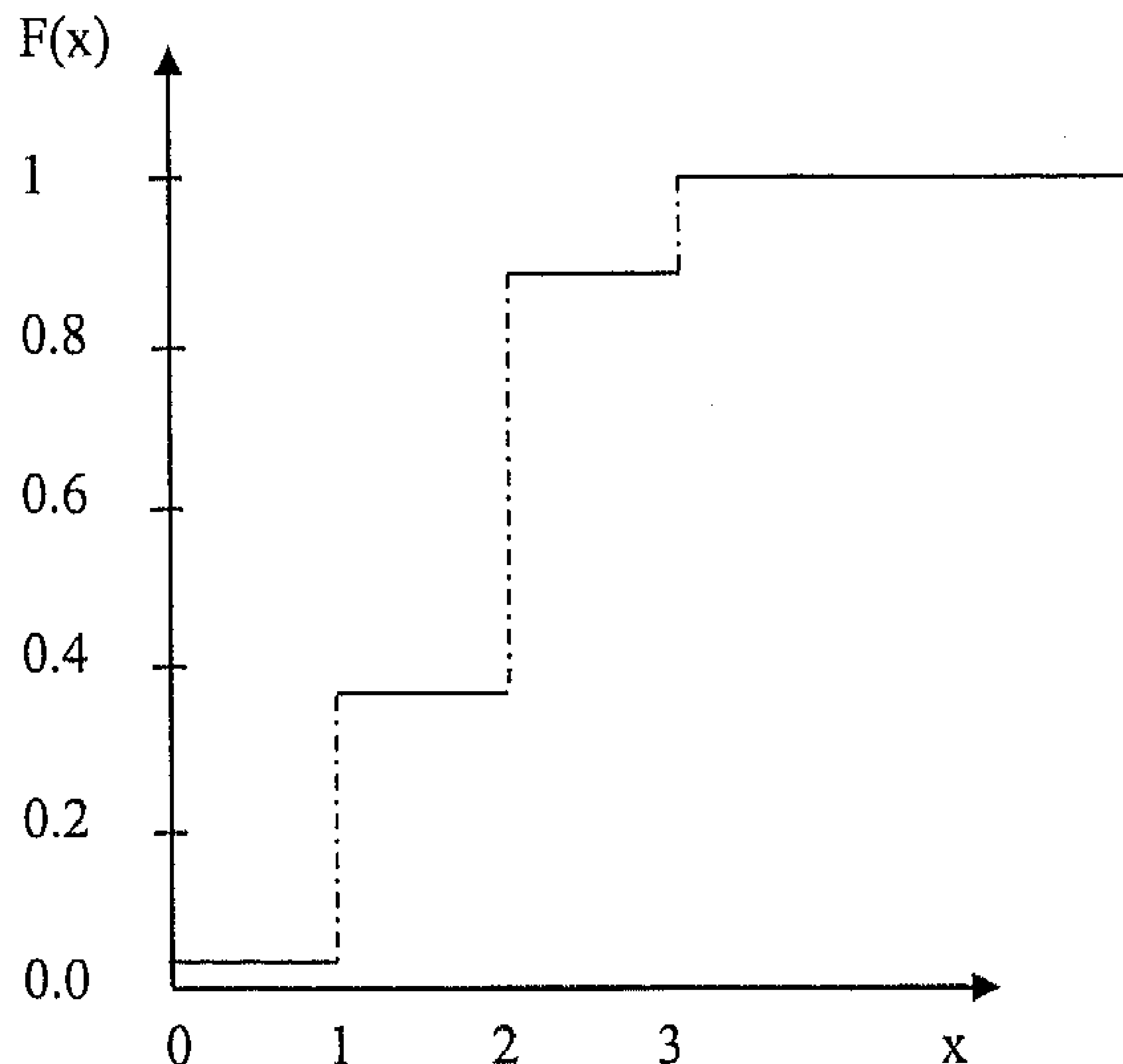
$$F(2) = P[X \leq 2] = p(0) + p(1) + p(2) = 0.029 + 0.343 + 0.514 = 0.886$$

$$F(3) = P[X \leq 3] = p(0) + p(1) + p(2) + p(3)$$
$$= 0.029 + 0.343 + 0.514 + 0.114 = 1$$

وسوف تكتب دالة التوزيع للمتغير X بالصورة التالية:

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.029, & 0 \leq x < 1 \\ 0.372, & 1 \leq x < 2 \\ 0.886, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

ويمكن تمثيل دالة التوزيع بيانيا كما يلي:



ونلاحظ من الرسم البياني أن دالة التوزيع هي دالة سُلمية غير تناقصية. وعموما نجد أن هناك علاقة بين التوزيع الاحتمالي ودالة التوزيع للمتغير العشوائي المتقطع وأن معرفة أحدهما يعطينا الأخرى؛ فبمعلومية دالة التوزيع يمكن تعيين دالة الكتلة الاحتمالية كالتالي:

$$p(x_k) = F(x_k) - F(x_{k-1}).$$

٧-٤-٣ المتغير العشوائي المتصل(المستمر) (Continuous Random Variable)

إذا كان X متغيراً عشوائياً متصلاً معرفاً في الفترة $-\infty < X < \infty$ فإنه توجد دالة $f(x)$ تسمى دالة الكثافة الاحتمالية والتي تحقق الشروط التالية:

(i) $f(x) \geq 0,$

(ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

خواص دالة الكثافة (Properties of Density Function):

١- إذا كانت A حادثة عشوائية بحيث $A = \{a < X < b\}$ فإن

$$P(A) = P[a < X < b] = \int_A f(x)dx = \int_a^b f(x)dx$$

٢- مجموعة الحوادث

$$\{a < X \leq b\}, \{a \leq X < b\}, \{a < X < b\}, \{a \leq X \leq b\}$$

لها نفس قيمة الاحتمال وهى

$$\int_a^b f(x)dx$$

٣- احتمال أن المتغير العشوائي المتصل X يأخذ قيمة مفردة يساوى صفراً بمعنى أن

$$P[X = a] = \int_a^a f(x)dx = 0$$

٤- دالة التوزيع للمتغير العشوائي المتصل تعرف من العلاقة التالية:

$$F(x) = P[X \leq x] = \int_{-\infty}^{x} f(u)du$$

ومن الخاصية السابقة نجد أنه يمكن تعيين دالة الكثافة الاحتمالية بدلالة دالة التوزيع كما يلي:

$$f(x) = \frac{d}{dx} F(x)$$

مثال (٧-٣)

إذا كان X متغيراً عشوائياً دالة كثافته الاحتمالية هي

$$f(x) = \begin{cases} k(1-x)x, & 0 < x < 1 \\ 0, & otherwise \end{cases}$$

(أ) أوجد قيمة الثابت k

(ب) أوجد قيمة $P[\ X\ >\ 0.2]$

(ج) أوجد دالة التوزيع التراكمي $F(x)$ لهذا المتغير

الحل

(أ) حيث إن $f(x)$ دالة كثافة احتمالية فإن

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\therefore \int_{-\infty}^{\infty} f(x)dx = \int_{0}^{1} k(1-x)x\, dx$$

$$= k\int_{0}^{1} (x - x^2)\, dx = k\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^1$$

$$= k\left[\frac{1}{2} - \frac{1}{3} - 0\right] = \frac{k}{6}$$

$$\Rightarrow \frac{k}{6} = 1 \Rightarrow k = 6$$

(ب) قيمة $P[\ X\ >\ 0.2]$

$$P[X > 0.2] = \int_{0.2}^{\infty} f(x)dx$$

$$= \int_{0.2}^{1} 6(1-x)x\,dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{0.2}^{1}$$

$$= 6\left[\left(\frac{1^2}{2} - \frac{1^3}{3}\right) - \left(\frac{0.2^2}{2} - \frac{0.2^3}{3}\right)\right] = 0.896$$

(ج) دالة التوزيع

- إذا كانت $x < 0$ فإن

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(u)du = \int_{-\infty}^{x} 0\,dx = 0$$

- إذا كانت $0 < x < 1$ فإن

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(u)du$$

$$= \int_{0}^{x} 6u(1-u)du = 6\left[\frac{u^2}{2} - \frac{u^3}{3}\right]_{0}^{x}$$

$$= 3x^2 - 2x^3 = x^2(3-2x)$$

- إذا كانت $x \ge 1$ فإن

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(u)du$$

$$= \int_{0}^{1} f(u)du + \int_{1}^{x} 0\,dx = 1$$

وبالتالي فإن دالة التوزيع يمكن كتابتها على الصورة التالية:

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2(3-2x), & 0 < x < 1 \\ 1, & x \ge 1 \end{cases}$$

٧-٥ التوقع الرياضي (Mathematical Expectation)

٧-٥-١ التوقع الرياضي للمتغير X

(أ) إذا كان X متغيراً عشوائياً منفصلاً له دالة الكتلة الاحتمالية $p(x)$ فإن القيمة المتوقعة للمتغير X يرمز لها بالرمز $E[X]$ وتعرف كالتالي:

$$E[X] = \sum_k x_k \, p(x_k)$$

(ب) إذا كان X متغيراً عشوائياً متصلاً(مستمراً) له دالة الكثافة الاحتمالية $f(x)$ فإن القيمة المتوقعة للمتغير X يرمز لها بالرمز $E[X]$ وتعرف كالتالي:

$$E[X] = \int_{-\infty}^{\infty} x \, f(x) dx$$

٧-٥-٢ التوقع الرياضي لأي دالة في X

إذا كان X متغيراً عشوائياً وكانت $g(X)$ دالة في X فإن القيمة المتوقعة لهذه الدالة هي

- إذا كان X متغيراً متقطعاً فإن

$$E[g(X)] = \sum_k g(x_k) \, p(x_k)$$

- إذا كان X متغيراً متصلاً فإن

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \, f(x) dx$$

مثال (٧-٤)

قام فصل دراسي يحتوي على 120 طالباً وطالبة برحلة دراسية مستقلين 3 حافلات. هناك 36 طالباً في إحدى الحافلات، 40 في الحافلة الثانية ، 44 في الحافلة الثالثة. عند وصول الحافلات، تم اختيار طالب عشوائيا من 120 طالبا. بفرض أن X يرمز لعدد الطلاب في الحافلة التي تم اختيار ذلك الطالب عشوائيا منها، أوجد E[X].

الحل:

حيث إن اختيار الطالب عشوائياً فإن جميع الطلاب متساوون في فرصة الاختيار وبالتالي فإن

$$P[X=36]=\frac{36}{120},\ P[X=40]=\frac{40}{120},\ P[X=44]=\frac{44}{120}$$

$$\therefore E[X]=\sum_{x} xp(x)$$

$$=36\times\left(\frac{36}{120}\right)+40\times\left(\frac{40}{120}\right)+44\times\left(\frac{44}{120}\right)=\frac{1208}{30}=40.267$$

مثال (٧-٥)

ألقيت قطعة عملة ثلاث مرات متتالية. وكان المتغير العشوائي X يمثل عدد مرات ظهور صورة. فأوجد $E(X)$، $E(X^2)$

الحل

فضاء العينة لهذه التجربة هو

$$S=\{(H,H,H),(H,H,T),(H,T,H),(H,T,T),(T,H,H),(T,H,T),(T,T,H),(T,T,T)\}$$

فإن القيم الممكنة للمتغير العشوائي X هي: 3 ,2, 1, 0

ويمكن تعيين دالة الكتلة الاحتمالية كالتالي:

$$p(0)=P[X=0]=P[(T,T,T)]=\frac{1}{8}=0.125$$

$$p(1)=P[X=1]=P[(H,T,T),(T,H,T),(T,T,H)]=\frac{3}{8}=0.375$$

$$p(2)=P[X=2]=P[(H,H,T),(H,T,H),(T,H,H)]=\frac{3}{8}=0.375$$

$$p(3)=P[X=3]=P[(H,H,H)]=\frac{1}{8}=0.125$$

ويمكن كتابتها في الجدول التالي

X	0	1	2	3	المجموع
$p(x)$	0.125	0.375	0.375	0.125	1

- من قانون التوقع الرياضي

$$E[X] = \sum_k x_k p(x_k)$$

X	0	1	2	3	المجموع
$p(x)$	0.125	0.375	0.375	0.125	1
$xp(x)$	0	0.375	0.75	0.375	1.5

ومن الجدول السابق فإن

$$E[X] = \sum_k x_k p(x_k) = 1.5$$

- إذا كانت $g(X) = X^2$ فإن

$$E[X^2] = \sum_k x_k^2 p(x_k)$$

X	0	1	2	3	المجموع
$p(x)$	0.125	0.375	0.375	0.125	1
$x^2\ p(x)$	0	0.375	1.5	1.125	3.0

ومن الجدول السابق فإن

$$E[X^2] = \sum_k x_k^2 p(x_k) = 3.0$$

مثال (٧-٦)

إذا كان X متغيراً متصلاً دالة كثافة الاحتمالية هي

$$f(x) = \begin{cases} \frac{1}{2} e^{-x}, & x \geq 0 \\ \frac{1}{2} e^{x}, & x < 0 \end{cases}$$

فإذا كانت $Y = |X|$ أوجد $E[Y]$

الحل

$$\because |X| = \begin{cases} x, & x \geq 0 \\ -x, & x < 0 \end{cases}$$

$$\therefore \mu = E[X] = \int_{-\infty}^{\infty} |x| f(x) dx = \int_{0}^{\infty} |x| f(x) dx + \int_{-\infty}^{0} |x| f(x) dx$$

$$= \int_{0}^{\infty} xf(x)dx - \int_{-\infty}^{0} xf(x)dx$$

$$= \frac{1}{2}\int_{0}^{\infty} x\, e^{-x}dx - \frac{1}{2}\int_{-\infty}^{0} x\, e^{x}dx$$

وبتعيين التكاملات السابقة بالتجزيء فنحصل على

$$\therefore E[X] = \frac{1}{2} + \frac{1}{2} = 1$$

خواص التوقع الرياضي:

١- التوقع الرياضي لمقدار ثابت هو نفس المقدار الثابت، أي أن

$$E[c] = c$$

حيث c مقدار ثابت.

٢- إذا كان X متغيراً عشوائياً فإن

$$E[aX] = a\,E[X]$$

٣- إذا كان X متغيراً عشوائياً a, b ثوابت فإن

$$E[aX + b] = aE[X] + b$$

٤- إذا كانت $g(X) = g_1(X) + g_2(X)$ فإن

$$E[g(X)] = E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$$

مثال (٧-٦)

في مثال(٧-٥)، أوجد قيمة $E[3X + 2]$

الحل:

من قانون التوقع فان

$$E[3X + 2] = \sum_{k} (3x_k + 2)\, p(x_k)$$

X	0	1	2	3	المجموع
$3X + 2$	2	5	8	11	
$p(x)$	0.125	0.375	0.375	0.125	1
$(3X + 2)p(x)$	0.25	1.875	3	1.375	6.5

ومن الجدول السابق فإن

$$E[3X+2]=\sum_k (3x_k+2)p(x_k)=6.5$$

ويمكن استخدام الخواص السابقة للتوقع لتعيين المطلوب كالتالي:

$$E[3X+2]=3E[X]+2=3(1.5)+2=6.5$$

٦-٧ التباين (Variance)

إذا كان X متغيراً عشوائياً له القيمة المتوقعة $E[X]$ فإن تباين هذا المتغير والذى يرمز له بالرمز Var(X) يعرف كالآتي:

$$Var(X)=E[(X-E[X])^2]$$

ويعرف الانحراف المعياري للمتغير العشوائي على أنه الجذر الموجب للتباين. وباستخدام خواص التوقع الرياضي يمكن الحصول على صيغة مختصرة للتباين كالتالي:

$$Var(X)=E[X^2]-(E[X])^2$$

الإثبات

$$\begin{aligned}Var(X)&=E[(X-E[X])^2]\\&=E[X^2-2XE[X]+(E[X])^2]\\&=E[X^2]-2(E[X])^2+(E[X])^2\\&=E[X^2]-(E[X])^2\end{aligned}$$

خواص التباين:

١- تباين المقدار الثابت يساوى صفراً

$$Var(c)=0$$

٢- إذا كان X متغيراً عشوائياً وكان a ثابت فإن

$$Var(aX)=a^2Var(X)$$

٣- إذا كان X متغيراً عشوائياً وكان b ثابت فإن

$$Var(X+b)=Var(X)$$

٤- إذا كان X متغيراً عشوائياً وكان a, b ثوابت فإن

$$Var(aX+b)=a^2Var(X)$$

مثال (٧-٧)

إذا كان X متغيراً عشوائيا يمثل عدد الصور التي ستظهر عند إلقاء قطعة عملة مرتين متتاليتين. فأوجد التوقع والتباين للمتغير X

الحل:

فضاء العينة لهذه التجربة هو

$$S=\{(H,H),(H,T),(T,H),(T,T)\}$$

المتغير X والذي يمثل عدد مرات ظهور صورة في تجربة إلقاء قطعة العملة مرتين يأخذ القيم 0, 1, 2 وله دالة الكتلة الاحتمالية التالية

X	0	1	2	المجموع
p(x)	0.25	0.5	0.25	1

- من قانون التوقع

$$E[X]=\sum_k x_k p(x_k)$$

X	0	1	2	المجموع
$p(x)$	0.25	0.5	0.25	1
$xp(x)$	0	0.5	0.5	1

ومن الجدول السابق فإن

$$E[X]=\sum_k x_k p(x_k)=1$$

وحيث ان

$$E[X^2]=\sum_k x_k^2 p(x_k)$$

X	0	1	2	المجموع
$p(x)$	0.25	0.50	0.25	1
$x^2\ p(x)$	0	0.50	1.00	1.5

ومن الجدول السابق فإن

$$E[X^2] = \sum_k x_k^2 p(x_k) = 1.5$$

- وبذلك يمكن تعيين التباين كالتالي

$$Var(X) = E[X^2] - (E[X])^2 = 1.5 - (1)^2 = 0.5$$

مثال (٧-٨)

إذا كان X متغيراً عشوائياً له دالة التوزيع

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.029, & 0 \le x < 1 \\ 0.372, & 1 \le x < 2 \\ 0.886, & 2 \le x < 3 \\ 1, & x \ge 3 \end{cases}$$

فأوجد

(أ) التوزيع الاحتمالي للمتغير X

(ب) قيم الاحتمالات $P[1 < X \le 3]$, $P[X > 2]$

(ج) قيمة التوقع والتباين.

الحل

(أ) من العلاقة بين دالة التوزيع والتوزيع الاحتمالي للمتغير العشوائي

$$p(x_k) = F(x_k) - F(x_{k-1})$$

ومن دالة التوزيع فإن قيم المتغير X هي 3 ,2 ,1 ,0 وبالتالي فإن

$$p(0) = P[X = 0] = F(0) = 0.029$$

$$p(1) = P[X = 1] = F(1) - F(0) = 0.372 - 0.029 = 0.343$$

$$p(2) = P[X = 2] = F(2) - F(1) = 0.886 - 0.372 = 0.514$$

$$p(3) = P[X = 3] = F(3) - F(2) = 1 - 0.886 = 0.114$$

وبالتالي فإن التوزيع الاحتمالي للمتغير X هو:

X	0	1	2	3	المجموع
$p(x)$	0.029	0.343	0.514	0.114	1

(ب) قيم الاحتمالات $P[1 < X \leq 3]$, $P[X > 2]$

$$P[X > 2] = P[X = 3] = 0.114$$

$$P[1 < X \leq 3] = F(3) - F(1) = 1 - 0.372 = 0.628$$

(ج) التوقع والتباين للمتغير X

من قانون التوقع الرياضي فإن:

$$E[X] = \sum_k x_k p(x_k)$$
$$= 0(0.029) + (1)(0.343) + (2)(0.514) + (3)(0.114) = 1.713$$

ومن قانون التباين فإن:

$$Var(X) = E[X^2] - (E[X])^2$$

$$E[X^2] = \sum_k x_k^2\, p(x_k)$$
$$= (0)^2(0.029) + (1)^2(0.343) + (2)^2(0.514) + (3)^2(0.114) = 3.425$$

$$\therefore Var(X) = 3.425 - (1.713)^2 = 0.4906$$

مثال (٧-٩)

إذا كان X متغيراً عشوائياً يأخذ القيم 1 ,0 ,1- فقط، وإذا كان

(أ) $P[X = 0] = \frac{1}{2}$ أوجد $E[X^2]$

(ب) $P[X = 0] = \frac{1}{2}, E[X] = \frac{1}{6}$ فأوجد $P[X = 1], P[X = -1]$

الحل

(أ) يمكن كتابة التوزيع الاحتمالي للمتغير X على الصورة

X	-1	0	1	المجموع
$p(x)$	a	1/2	b	1

حيث a, b مقادير موجبة بحيث

$$\sum_k p(x_k) = 1$$

$$\Rightarrow a + b + \frac{1}{2} = 1$$

$$\Rightarrow a + b = \frac{1}{2} \qquad (i)$$

لكن

$$E[X^2] = \sum_k x_k^2 p(x_k)$$

$$= (-1)^2 a + (0)^2 \left(\frac{1}{2}\right) + (1)^2 b = a + b = \frac{1}{2}$$

(ب) إذا كانت $E[X] = \frac{1}{6}$, $P[X = 0] = \frac{1}{2}$ فإن

$$E[X] = \sum_k x_k p(x_k) = (-1)a + (0)\left(\frac{1}{2}\right) + (1)b = \frac{1}{6}$$

$$b - a = \frac{1}{6} \qquad (ii)$$

من $(i), (ii)$ ينتج أن

$$a = \frac{1}{6}, \quad b = \frac{2}{6}$$

$$\Rightarrow P[X = -1] = \frac{1}{6}, \quad P[X = 1] = \frac{2}{6}$$

مثال (٧-١٠)

إذا كان X متغيراً عشوائياً له دالة الكثافة الاحتمالية

$$f(x) = \begin{cases} c(1 - x^2), & -1 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

أوجد

(أ) قيمة الثابت c

(ب) دالة التوزيع.

(ج) قيمة $P[0 < X < \frac{3}{4}]$

(د) قيمة التوقع الرياضي والتباين للمتغير X

الحل

(أ) حيث إن $f(x)$ دالة كثافة احتمالية فإن

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\therefore \int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{1} c(1-x^2)\,dx$$

$$= c\int_{-1}^{1}(1-x^2)\,dx = c\left[x-\frac{x^3}{3}\right]_{-1}^{1}$$

$$= c\left[\left(1-\frac{1}{3}\right)-\left(-1+\frac{1}{3}\right)\right] = \frac{4c}{3}$$

$$\Rightarrow \frac{4c}{3} = 1 \Rightarrow c = \frac{3}{4}$$

(ب) دالة التوزيع

- إذا كانت $x < -1$ فإن

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(u)du = \int_{-\infty}^{x} 0\,dx = 0$$

- إذا كانت $-1 < x < 1$ فإن

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(u)du$$

$$= \int_{-1}^{x}\frac{3}{4}(1-u^2)du = \frac{3}{4}\left[u-\frac{u^3}{3}\right]_{-1}^{x}$$

$$= \frac{3}{4}\left[x-\frac{x^3}{3}-\left(-1+\frac{1}{3}\right)\right] = \frac{1}{4}(2+3x-x^3)$$

- إذا كانت $x \ge 1$ فإن

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(u)du$$

$$= \int_0^1 f(u)du + \int_1^x 0\, dx = 1$$

دالة التوزيع يمكن كتابتها على الصورة التالية:

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{4}(2 + 3x - x^3), & -1 < x < 1 \\ 1, & x \geq 1 \end{cases}$$

(ج) قيمة الاحتمال

$$P\left[0 < X < \frac{3}{4}\right] = \int_0^{0.75} f(x)dx$$

$$= \int_0^{0.75} \frac{3}{4}(1 - x^2)\, dx = \frac{3}{4}\left[x - \frac{x^3}{3}\right]_0^{0.75}$$

$$= \frac{3}{4}\left[0.75 - \frac{1}{3}(0.75)^3 - 0\right] = 0.457$$

ويمكن استخدام دالة التوزيع في تعيين الاحتمال السابق كما يلي:

$$P\left[0 < X < \frac{3}{4}\right] = F\left(\frac{3}{4}\right) - F(0)$$

$$= \frac{1}{4}\left(2 + \frac{9}{4} - \frac{27}{64}\right) - \frac{1}{4}(2 + 0) = 0.457$$

(د) التوقع والتباين

أولاً: تعيين قيمة التوقع كما يلي:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

$$= \int_{-1}^{1} \frac{3}{4}(x - x^3)dx = 2\int_0^1 \frac{3}{4}(x - x^3)dx$$

$$= \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)_0^1 = \frac{3}{2}\left(\frac{1}{2} - \frac{1}{4} - 0\right) = \frac{3}{8}$$

ثانياً: تعيين قيمة التباين كما يلي:

بحساب قيمة

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx$$

$$= \int_{-1}^{1} \frac{3}{4}(x^2 - x^4)dx = 2\int_{0}^{1} \frac{3}{4}(x^2 - x^4)dx$$

$$= \frac{3}{2}\left(\frac{x^3}{3} - \frac{x^5}{5}\right)_0^1 = \frac{3}{2}\left(\frac{1}{3} - \frac{1}{5} - 0\right) = \frac{1}{5}$$

فإن التباين

$$Var(X) = E[X^2] - (E[X])^2 = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = 0.0594$$

أسئلة وتمارين (٧)

١- عرف المتغير العشوائي مع ذكر أنواعه.

٢- اذكر خواص دالة التوزيع الاحتمالية.

٣- ما هي العلاقة بين كلٍّ من دالة التوزيع ودالة الكثافة الاحتمالية؟

٤- ما المقصود بوصف المتغير العشوائي؟

٥- اذكر خواص دالة التوزيع التراكمية.

٦- في تجربة إلقاء حجر نرد مرتين متتاليتين. أوجد القيم الممكنة للمتغيرات العشوائية التالية:

(أ) القيمة العظمى لما سيظهر في الرميتين.

(ب) أقل قيمة ستظهر في الرميتين.

(ج) مجموع ما سيظهر في الرميتين.

(د) الفرق بين ما سيظهر في الرمية الأولى والرمية الثانية.

٧- في التمرين السابق أوجد التوزيع الاحتمالي للمتغيرات في (أ)، (ب)، (ج)، (د).

٨- إذا كان X متغير عشوائي دالة كثافة احتماله هي

$$f(x) = \frac{1}{2} e^{-|x|}, -\infty < x < \infty$$

(أ) اثبت أن هذه الدالة تمثل دالة كثافة احتمالية.

(ب) أوجد $P[X > 3], \quad P[-2 < X < 5]$

(ج) أوجد التوقع الرياضي والتباين للمتغير العشوائي X

٩- إذا كان X متغيراً عشوائياً له دالة التوزيع الآتية:

$$F(x) = \begin{cases} 1 - \frac{1}{x}, & x > 1 \\ 0, & x < 1 \end{cases}$$

(أ) أوجد دالة الكثافة، وبين عند أي نقطة تكون $f(x)$ غير معرفة.

(ب) احسب قيمة الاحتمالات $P[X < 2], P[5 < X < 10]$ من الدوال $F(x), f(x)$

١٠- إذا كانت دالة التوزيع الاحتمالي للمتغير X هي

X	1	2	3	4
$p(x)$	c	c/2	c/3	c/4

(أ) أوجد قيمة الثابت c

(ب) أوجد قيمة $P[X \geq 2]$

(ج) احسب قيمة التوقع الرياضي والتباين للمتغير X

١١- رُمي حجرا نرد معاً وكان X يمثل مجموع ما سيظهر على الحجرين

(أ) كون جدول التوزيع الاحتمالي للمتغير X

(ب) أوجد دالة التوزيع للمتغير X

١٢- إذا كان Y متغيراً عشوائياً دالة كثافته الاحتمالية هي

$$g(y) = \begin{cases} cy^2, -1 < y < 1 \\ 0, \quad \text{otherwise} \end{cases}$$

(أ) أوجد قيمة الثابت c

(ب) أوجد دالة التوزيع $G(y)$

(ج) احسب قيمة الاحتمالات

$$P[0 < Y < 1], P[0 < Y \leq 3], P[Y = 0.5], P[-1 < Y \leq 0.5]$$

(د) أوجد قيمة التباين للمتغير العشوائي Y

١٣- إذا كان X متغيراً عشوائياً له دالة التوزيع

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

(أ) احسب قيمة الاحتمالات $P[X \leq 0.5]$, $P[0.25 < X < 0.75]$

(ب) أوجد قيمة التوقع الرياضي والتباين.

١٤- إذا كان X متغيراً عشوائياً له التوزيع الاحتمالي

X	0	1	2	a	6
$p(x)$	0.1	0.1	0.3	0.2	0.3

أوجد قيمة a عندما

(أ) $E[X] = 3$

(ب) $E[X^2] = 18$

١٥- إذا كان X متغيراً عشوائياً متصلاً له دالة الكثافة

$$f(x) = \begin{cases} 3x^a, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

(أ) أوجد قيمة a

(ب) احسب التوقع الرياضي للمتغير X والتباين.

(ج) أوجد الاحتمال $P[0.25 < X < 0.5]$

١٦- إذا علم أنه بالنسبة للمتغير العشوائي X تتحقق العلاقتين

$$E[X + 4] = 10\,, \qquad E[(X+4)^2] = 116$$

فأوجد قيمة التوقع الرياضي والتباين للمتغير X

١٧- إذا كان X متغيراً عشوائياً متقطعاً له التوزيع الاحتمالي التالي:

$$P[X = x] = \frac{c}{x+2}, \quad x = 0, 1, 2, 3$$

(أ) أوجد قيم الثابت c

(ب) باستخدام الرموز $E_1 = \{X = 1\}, E_2 = \{X = 2\}, E_3 = \{X = 3\}$ أوجد قيم الاحتمالات $P(E_3 \cup E_1),\ P(E_1^c)$

١٨- إذا كانت دالة التوزيع التراكمية للمتغير العشوائي X على الصورة التالية:

$$F(x) = \begin{cases} 0, & x < -2 \\ 0.2, & -2 \le x < 2 \\ 0.5, & 2 \le x < 3 \\ c, & 3 \le x < 5 \\ 1, & x \ge 5 \end{cases}$$

فأوجد

(أ) قيمة الثابت c إذا علمت أن $p(3) = 0.25$

(ب) احتمالات الحوادث التالية:

$$\{1 < X < 2\}, \{X > 1.5\}, \{0.5 < X < 2.5\}$$

١٩- إذا كان لدينا صندوقان يحتوى الأول 4 كرات بيضاء وكرتين سوداوين بينما يحتوى الثاني 5 كرات بيضاء، 3 كرات سوداء. اخترنا عشوائياً كرة من الصندوق الأول ووضعناها في الصندوق الثاني، ثم اخترنا من الصندوق الثاني عينه من 3 كرات. إذا عرفنا المتغير X بحيث يمثل عدد الكرات البيضاء في العينة فاحسب

(أ) دالة الكتلة الاحتمالية للمتغير X

(ب) التوقع والتباين للمتغير X

(ج) التوقع والتباين للمتغير Y بحيث $Y = 3X + 2$

٢٠- إذا تم اختيار ثلاث كرات عشوائياً من وعاء يحتوى على 15 كرة مرقمة من 1 حتى 15 وإذا كان هدفنا الحصول على كرة من الكرات المسحوبة على الأقل عليها رقم أكبر من أو يساوى 13 فأوجد التوزيع الاحتمالي للحصول على الهدف المرجو من التجربة.

٢١- إذا كانت دالة الكتلة الاحتمالية للمتغير X تعطى بالصورة

$$p(i) = \frac{c\,\lambda^i}{i!},\;\; i = 0, 1, 2, \dots, \lambda > 0$$

فأوجد

(أ) قيمة الثابت c

(ب) الاحتمالات $P[X = 0],\;\; P[X > 2]$

٢٢- أوجد $E[X]$ إذا كان X يمثل الرقم الظاهر عند إلقاء حجر نرد مرة واحدة؟

٢٣- فصل دراسي يحتوى علي 150 طالباً قام برحلة ترفيهية حيث تم تقسيم الطلاب على أربعة أتوبيسات بحيث الأتوبيس الأول يحمل 40 طالباً، الأتوبيس الثاني يحمل 43 طالباً، الأتوبيس الثالث يحمل 32 طالباً، والأتوبيس الرابع يحمل 35 طالباً. عند وصول الأتوبيسات سيتم اختيار طالب عشوائياً من 150 طالباً ليكون منسق الرحلة. بفرض أن X يمثل عدد الطلاب في الأتوبيس الذي تم اختيار الطالب عشوائياً منه، فأوجد $E[X]$ وايضا $Var(X)$

٢٤- بفرض أن X يعبر عن الفرق بين عدد الصور وعدد الكتابات عند إلقاء قطعة عمله n من المرات. فما هي القيم الممكنة للمتغير X؟

٢٥- إذا كانت دالة التوزيع للمتغير X معطاة كالتالي:

$$F(x)=\begin{cases}0, & x<0\\ \dfrac{x}{4}, & 0\le x<1\\ \dfrac{1}{2}+\dfrac{x-1}{4}, & 1\le x<2\\ \dfrac{11}{12}, & 2\le x<3\\ 1, & x\ge 3\end{cases}$$

فأوجد

(أ) التوزيع الاحتمالي للمتغير X

(ب) الاحتمالات: $P[X>1]$, $P[X\le 2]$, $P[0.5<X<1.5]$

٢٦- من صندوق يحتوى 15 عنصراً تم اختيار عينة من 3 عناصر عشوائياً. فإذا كان يحتوى الصندوق على 5 عناصر تالفة. أوجد العدد المتوقع من العناصر التالفة في العينة المختارة من الصندوق.

٢٧- إذا كان لدينا صندوق يحتوى على 5 كرات حمراء، 5 كرات زرقاء. قام شخص باختيار كرتين عشوائياً فإذا كانت الكرتين من نفس اللون سيكسب 1.50 ريالاً سعودياً، وإذا كانت ألوان الكرتين مختلفين سيخسر الشخص ريال سعودي واحد فأوجد

(أ) القيمة المتوقعة لمكسب هذا الشخص.

(ب)التباين لمكسب الشخص.

٢٨- إذا كانت دالة الكثافة الاحتمالية للمتغير X الذي يمثل زمن حياة نوع معين من الأجهزة الالكترونية (مُقاس بالساعة) تعطى بالصورة التالية:

$$f(x)=\begin{cases}\dfrac{10}{x^2}, & x>10\\ 0, & x\le 10\end{cases}$$

فأوجد

(أ) احتمال $P[X>20]$

(ب) دالة التوزيع التراكمية للمتغير X

(ج) أوجد احتمال أنه من بين 6 أجهزة سيعمل على الأقل 3 أجهزة منها على الأقل 15 ساعة.

٢٩- إذا كان زمن حياة جهاز الكمبيوتر بالساعات متغيراً عشوائياً له دالة الكثافة الاحتمالية التالية:

$$f(x) = x\,e^{-x},\ \ x \geq 0$$

فأوجد التوقع الرياضي والتباين لزمن حياة جهاز الكمبيوتر.

٣٠- إذا كانت دالة الكثافة الاحتمالية للمتغير X تعطى بالعلاقة

$$f(x) = \begin{cases} cx^n, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

فأوجد

(أ) قيمة الثابت c

(ب) $E[X],\ Var(X)$

(ج) دالة التوزيع التراكمية للمتغير X

٣١- بفرض أن X يعبر عن الفرق بين ما سيظهر على حجري النرد عند إلقاء كلٍّ منهما مرة واحدة. فأوجد

(أ) التوزيع الاحتمالي للمتغير X

(ب)دالة التوزيع للمتغير X

(ج) التوقع والتباين للمتغير X

الفصل الثامن

بعض التوزيعات الاحتمالية المنفصلة

SOME DISCRETE PROBABILITY DISTRIBUTIONS

٨–١ مقدمة

تم التعرض في الفصل السابق لمفهوم المتغير العشوائي وأنواعه كما تم التعرض للمفاهيم الخاصة للمتغيرات وتعريف التوقع الرياضي والتباين. قد يكون المتغير العشوائي يتبع توزيعاً احتمالياً معيناً بمعنى أن دالة كتلته الاحتمالية أو دالة كثافته لها شكل معين. تنقسم التوزيعات الاحتمالية إلى نوعين تبعاً لنوع المتغير التي تصفه، فالتوزيعات الاحتمالية إما متقطعة (منفصلة) أو متصلة (مستمرة). وسندرس في هذا الفصل بعض التوزيعات المتقطعة وسنخصص الفصل التالي لبعض التوزيعات الاحتمالية المتصلة.

٨–٢ توزيع برنوللي (Bernoulli Distribution)

توزيع برنوللي من أبسط التوزيعات المنفصلة ويعتبر حالة خاصة من عديد من التوزيعات المنفصلة ويستخدم هذا التوزيع إذا كان هناك تجربة أو محاولة عشوائية تحقق الشروط التالية:

(أ) يمكن إجراء هذه التجربة مرة واحدة فقط.

(ب) لها نتيجتان فقط هما ظهور حدث معين وتسمى النتيجة نجاحاً أو عدم ظهوره وتسمى فشلاً.

(ج) احتمال النجاح مقدار ثابت واحتمال الفشل مقدار ثابت بحيث

$$p = P\left[\text{النجاح}\right]$$

$$q = P\left[\text{الفشل}\right]$$

حيث

$$p + q = 1$$

إذا عرفنا متغيراً عشوائياً X يمثل عدد مرات النجاح على فراغ العينة فإن:

$x = 1$ إذا حصلنا على النجاح (اذا ظهر الحدث المعين)، $x = 0$ إذا حصلنا على فشل (اذا لم يظهر الحدث المعين) ويمكن كتابة دالة التوزيع الاحتمالي للمتغير العشوائي X على الصورة الآتية:

$$p(x) = p^x q^{1-x}, \qquad x = 0, 1$$

حيث p هو احتمال النجاح، q هو احتمال الفشل، $q = 1 - p$

خواص توزيع برنوللي:

$1 - \sum_x p(x) = 1$

الإثبات:

$$\sum_x p(x) = \sum_{x=0}^{1} p^x q^{1-x} = q + p = 1$$

$2 - \mu = p$

الإثبات:

$$\mu = E[X] = \sum_x x\, p(x) = \sum_{x=0}^{1} x p^x q^{1-x} = 0(q) + 1(p) = p$$

$3 - \sigma^2 = pq$

الإثبات:

$$\because \sigma^2 = Var(X) = E[X^2] - \mu^2$$

حيث

$$E[X^2] = \sum_x x^2\, p(x) = \sum_{x=0}^{1} x^2 p^x q^{1-x}$$

$$= (0)^2 \times (q) + (1)^2 \times (p) = p$$

$$\Longrightarrow \sigma^2 = Var(X) = p - p^2 = p(1-p) = pq$$

وبالتالي فإن الانحراف المعياري للمتغير X هو

$$\sigma = \sqrt{pq}$$

٤- دالة التوزيع للمتغير العشوائي X الذي يتبع توزيع برنوللي تعطى من العلاقة التالية:

$$F(x) = P[X \le x] = \begin{cases} 0, & x < 0 \\ q, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

٨-٣ توزيع ذي الحدين (التوزيع الثنائي) (Binomial Distribution)

يعتبر توزيع ذات الحدين من أهم التوزيعات الإحصائية لمتغير عشوائي منفصل، فهو يتناول التجارب ذات الأهمية التطبيقية في العلوم المختلفة والتي يترتب عليها إحدى نتيجتين فقط، هما ظهور حدث معين (نجاح)، أو عدم ظهور هذا الحدث (فشل). ويمثل توزيع ذي الحدين تعميماً لمحاولة برنوللي بحيث نجد أن تجربة ذي الحدين تحقق الشروط التالية:

(أ) يمكن تكرارها n من المرات حيث $n \ge 1$،

(ب) كل محاولة لها نتيجتين فقط نجاح أو فشل،

(ج) احتمال النجاح والفشل مقدار ثابت حيث يرمز لاحتمال النجاح بالرمز p واحتمال الفشل بالرمز q،

(د) نتيجة كل محاولة مستقلة عن الأخرى.

وبفرض أن X يمثل عدد مرات النجاح خلال تجربة ذي الحدين فإن المتغير العشوائي X يتبع توزيع ذي الحدين بمعالم n, p وتعطى دالة كتلته الاحتمالية بالصورة الآتية:

$$p(x) = P[X = x] = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \dots, n, \qquad p + q = 1$$

حيث: n عدد مرات إجراء التجربة (عدد المحاولات)،

p احتمال نجاح التجربة (احتمال ظهور حدث معين) حيث $0 \le p \le 1$.

وأيضا $q = 1 - p$ احتمال فشل التجربة (احتمال عدم ظهور الحدث المعين) كما أن

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}, \quad n! = n(n-1)(n-2)\dots(2)(1),$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

فإذا كان X متغيراً عشوائياً يتبع توزيع ذي الحدين بمعالم n, p فإنه يكتب اختصاراً $X \sim Bin(n,p)$

خواص توزيع ذي الحدين:

$1 - \sum_x p(x) = 1$

الإثبات:

$$\sum_x p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p+q)^n = (1)^n = 1$$

$2 - \mu = np$

الإثبات:

$$\mu = E[X] = \sum_x x p(x) = \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=0}^{n} \frac{n!}{x!\,(n-x)!}\, x\, p^x q^{n-x}$$

$$= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-x)!}\; p^{x-1} q^{n-x}$$

$$= np \sum_{x-1=0}^{n} \binom{n-1}{x-1}\; p^{x-1} q^{(n-1)-(x-1)}$$

بوضع $t = x - 1$ سوف نحصل على

$$\mu = E[X] = np \sum_{t=0}^{n-1} \binom{n-1}{t}\; p^t q^{(n-1)-t}$$

$$= np(p+q)^{n-1} = np(1)^{n-1} = np$$

$3 - \sigma^2 = npq$

الإثبات:

$$\because\ \sigma^2 = Var(X) = E[X^2] - \mu^2$$

$$E[X^2] = E[X(X-1) + X]$$

$$= E[X(X-1)] + np \qquad (i)$$

$$E[X(X-1)] = \sum_x x(x-1)p(x)$$

$$= \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x q^{n-x}$$

$$= \sum_{x=2}^{n} \frac{n!}{x!\,(n-x)!}\ x(x-1)\ p^x q^{n-x}$$

$$= n(n-1)p^2 \sum_{x=2}^{n-2} \frac{(n-2)!}{(x-2)!\,(n-x)!}\ p^{x-2} q^{n-x}$$

$$= n(n-1)p^2 \sum_{x-2=0}^{n-2} \binom{n-2}{x-2}\ p^{x-2} q^{(n-2)-(x-2)}$$

بوضع $r = x - 2$ سوف نحصل على

$$E[X(X-1)] = n(n-1)p^2 \sum_{r=0}^{n-2} \binom{n-2}{r}\ p^r q^{(n-r)-r}$$

$$= n(n-1)p^2(p+q)^{n-2} = n(n-1)p^2(1)^{n-2}$$

$$= n(n-1)p^2 \qquad (ii)$$

بالتعويض من (ii) في (i) نحصل على

$$E[X^2] = n(n-1)p^2 + np$$

$$\therefore Var(X) = n(n-1)p^2 + np - (np)^2 = npq$$

ومن الواضح أن الانحراف المعياري لتوزيع ذي الحدين هو

$$\sigma = \sqrt{npq}$$

ملحوظة (٨-١)

في توزيع ذي الحدين بمعالم n, p إذا كانت $n = 1$ سوف نحصل على توزيع برنوللى بمعلم p.

مثال (٨-١)

إذا كانت نسبة التدخين في أحد المجتمعات هي 0.6. أخذت عينة مكونه من 20 شخصاً. احسب الاحتمالات التالية:

(أ) عدد المدخنين في هذه العينة هو 2

(ب) ولا شخص يدخن

(ج) على الأكثر شخصان يدخنان

الحل

من الواضح أن أي شخص له نتيجتان فقط إما مدخن أو غير مدخن ونتيجة كل شخص مستقلة عن الآخر ويوجد لدينا أكثر من شخص 20 شخصاً.

بفرض أن X متغير عشوائي يرمز لعدد المدخنين في العينة، n عدد عناصر العينة، p احتمال أن يكون الشخص مدخناً، q احتمال أن يكون الشخص غير مدخن حيث

$$n = 20, p = 0.6, q = 1 - p = 0.4$$

وبالتالي فإن X له توزيع ذي الحدين بحيث

$$p(x) = P[X = x] = \binom{20}{x}(0.6)^x(0.4)^{20-x}, \ \ x = 0, 1, \dots, 20$$

(أ) احتمال أن يكون عدد المدخنين 2 يساوي

$$P[X = 2] = \binom{20}{2}(0.6)^2(0.4)^{20-2} = 4.7 \times 10^{-6}$$

(ب) احتمال أن لا يوجد مدخنون هو

$$P[X = 0] = \binom{20}{0}(0.6)^0(0.4)^{20-0} = 10.995 \times 10^{-9}$$

(ج) احتمال أن يوجد مدخنين اثنين على الأكثر يساوي

$$P[X \leq 2] = \sum_{x=0}^{2} \binom{20}{x} (0.6)^x (0.4)^{20-x}$$

$$= \binom{20}{0} (0.6)^0 (0.4)^{20-0} + \binom{20}{1} (0.6)^1 (0.4)^{20-1}$$

$$+ \binom{20}{2} (0.6)^2 (0.4)^{20-2} = 5.0413 \times 10^{-6}$$

مثال (٨-٢)

إذا كان احتمال فوز فريق ما في مبارة لكرة القدم في الدوري هو 0.3 وأتيحت له فرصة اللعب 10 مباريات خلال الدور الثاني للدوري ويجب أن يفوز فريق في نهاية المباراة. فأوجد

(أ) احتمال أن يفوز مرتين على الأقل.

(ب) العدد المتوقع لمرات الفوز خلال الدوري.

(ج) قيمة التباين.

الحل

من الواضح أن عملية اللعب لها نتيجتان فقط، إما أن يفوز الفريق (نجاح) أو يخسر (فشل)، وسوف تتكرر عملية اللعب 10 مرات، ونتيجة المبارة في إحدى المباريات لا تؤثر على المباريات الأخرى. فإن عملية لعب 10 مباريات لكرة القدم تمثل تجربة ذي الحدين.

بفرض أن X متغير عشوائي يمثل عدد مرات الفوز، n عدد المباريات، p احتمال أن يفوز الفريق، q احتمال أن يخسر الفريق حيث

$$n = 10, \quad p = 0.3, \quad q = 1 - p = 0.7$$

وبالتالى فإن X له توزيع ذي الحدين وتكون دالة كتلته الاحتمالية على الصورة التالية:

$$p(x) = P[X = x] = \binom{10}{x} (0.3)^x (0.7)^{10-x}, \quad x = 0, 1, \dots, 10$$

(أ) احتمال أن يفوز الفريق مرتين على الأقل هو

$$P[X \geq 2] = 1 - P[X < 2]$$

$$= 1 - \{ P[X = 0] + P[X = 1]\}$$

$$= 1 - \left\{\binom{10}{0} (0.3)^0 (0.7)^{10-0} + \binom{10}{1} (0.3)^1 (0.7)^{10-1}\right\}$$

$$= 1 - \{(0.7)^{10} + 3(0.7)^{9}\} = 0.8507$$

(ب) العدد المتوقع لمرات الفوز

$$\mu = E[X] = np = 10 \times 0.3 = 3$$

(ج) قيمة التباين

$$\sigma^2 = npq = 10 \times 0.3 \times 0.7 = 2.1$$

مثال (٨-٣)

إذا كان X متغيراً عشوائياً يتبع توزيع ذات الحدين بمعالم n, p وكان $\mu = 10, \sigma^2 = 6$. فأوجد قيمة n, p

الحل

بما أن X له توزيع ذي الحدين فإن

$$\mu = E[X] = np = 10 \qquad (i)$$

$$\sigma^2 = Var(X) = npq = 6 \qquad (ii)$$

بالتعويض من العلاقة (i) في العلاقة (ii) نجد أن

$$10q = 6 \Rightarrow q = \frac{6}{10} = 0.6$$

$$\therefore q = 1 - p \Rightarrow p = 1 - 0.6 = 0.4$$

بالتعويض عن p فى العلاقة (i) نحصل على قيمة n وهى

$$n = 25$$

٨-٤ توزيع بواسون (Poisson Distribution)

توزيع بواسون له تطبيقات كثيرة في الحياة العلمية والعملية. فهو يقدم بصورة خاصة نموذجاً جيداً للمعلومات التي تأخذ شكل التعداد. فهو بجانب أنه يصف الظواهر النادرة الحدوث مثل عدد حوادث السير خلال زمن محدد، عدد الأخطاء المطبعية في صفحة ما، عدد الزلازل السنوية. فإنه يصف كثيراً من الظواهر التي تحدث في الزمن أو الفراغ مثل عدد الذرات في ميكروثانية عن كمية من مادة مشعة، عدد المكالمات التي تصل سنترال ما في فترة زمنية معينة، عدد السلع التالفة التي ينتجها مصنع ما في فترة زمنية معينة.

وبفرض أن X يمثل عدد مرات النجاح التي تحدث في فترة زمنية معينة أو مكان محدد بمعدل λ فإن المتغير العشوائي X يتبع توزيع بواسون بمعلمة λ ودالة كتلته الاحتمالية تعطى بالصورة التالية:

$$p(x) = P[X = x] = \frac{e^{-\lambda}\lambda^{x}}{x!}, \qquad x = 0, 1, 2, ...$$

حيث e مقدار ثابت ويسمى الأساس الطبيعي للوغاريتمات ومقداره 2.7183 وللاختصار فيكتب المتغير العشوائي X الذي يتبع توزيع بواسون بمعلمة λ بالصورة التالية $X \sim Poisson(\lambda)$.

خواص توزيع بواسون:

$$1 - \sum_{x} p(x) = 1$$

الإثبات:

$$\sum_{x} p(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!} = e^{-\lambda}\, e^{\lambda} = e^{0} = 1$$

حيث

$$e^{\theta} = 1 + \frac{\theta}{1!} + \frac{\theta^{2}}{2!} + \frac{\theta^{3}}{3!} + \ ... + \frac{\theta^{i}}{i!} + \cdots = \sum_{k=0}^{\infty} \frac{\theta^{k}}{k!}$$

$$2 - \mu = \lambda$$

الإثبات:

$$\mu = E[X] = \sum_{x} xp(x) = \sum_{x=0}^{\infty} x\, \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x}}{(x-1)!} = \lambda\, e^{-\lambda} \sum_{x-1=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

بوضع $k = x - 1$ نحصل على

$$\mu = E[X] = \lambda\, e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!} = \lambda\, e^{-\lambda}\, e^{\lambda} = \lambda\, e^{0} = \lambda$$

$$3 - \sigma^{2} = \lambda$$

الإثبات:

$$\because\ \sigma^{2} = Var(X) = E[X^{2}] - \mu^{2}$$

$$E[X^2] = E[X(X-1)+X] = E[X(X-1)] + E[X]$$
$$= E[X(X-1)] + \lambda \qquad (i)$$

$$E[X(X-1)] = \sum_x x(x-1)p(x) = \sum_{x=0}^{\infty} x(x-1)\,\frac{e^{-\lambda}\lambda^x}{x!}$$
$$= e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^x}{(x-2)!} = \lambda^2\, e^{-\lambda}\sum_{x-2=0}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}$$

بوضع $r = x - 2$ نحصل على

$$E[X(X-1)] = \lambda^2\, e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda^2\, e^{-\lambda}\, e^{\lambda} = \lambda^2\, e^0 = \lambda^2 \quad (ii)$$

بالتعويض من (ii) في (i) فنحصل على

$$E[X^2] = \lambda^2 + \lambda$$
$$\Longrightarrow \sigma^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

نظرية (٨-١)

في توزيع ذي الحدين عندما تكون n كبيرة واحتمال النجاح مقدار صغير، فإن توزيع ذي الحدين بمعالم n, p يمكن تقريبه بتوزيع بواسون بمعلمة $\lambda = np$

البرهان

باستخدام دالة التوزيع الاحتمالي لتوزيع ذي الحدين فإن

$$Bin(n,p) = \binom{n}{x} p^x q^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}$$
$$= \frac{n!}{x!\,(n-x)!}\, p^x (1-p)^{n-x}$$

وبفرض $\lambda = np$ فإن $p = \frac{\lambda}{n}$ وبالتعويض فى دالة الكتله الاحتمالية لتوزيع ذى الحدين فإن:

$$Bin(n,p) = \frac{n!}{x!\,(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$=\frac{\lambda^x}{x!}\frac{n(n-1)(n-2)\dots(n-x+1)}{n^x}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$=\frac{\lambda^x}{x!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{x-1}{n}\right)\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-x}$$

بأخذ نهاية الطرفين عندما تكون n كبيرة أي عندما تقترب من اللانهاية نجد أن:

$$\lim_{n\to\infty} Bin(n,p)=\frac{\lambda^x}{x!}\left(1-\frac{1}{\infty}\right)\left(1-\frac{2}{\infty}\right)\dots\left(1-\frac{x-1}{\infty}\right)\left(1-\frac{\lambda}{\infty}\right)^{-x}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n$$

$$=\frac{\lambda^x}{x!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=\frac{e^{-\lambda}\lambda^x}{x!}=Poisson(\lambda)$$

في توزيع ذي الحدين إذا كانت عدد المحاولات $n > 30$ وكذلك $np < 5, n(1-p) < 5$ فإننا سوف نستخدم توزيع بواسون بدلاً من توزيع ذي الحدين بحيث $\lambda = np$

مثال (٨-٤)

إذا كان احتمال وجود شخص أعسر (يستخدم يده اليسرى) في مجتمع ما هو 0.01 تم اختيار عينة من 400 شخص عشوائياً من ذلك المجتمع. فأوجد

(أ) احتمال وجود 4 على الأقل يستخدمون اليد اليسرى في هذه العينة.

(ب) العدد المتوقع من هؤلاء الأشخاص في العينة.

الحل

بفرض أن X يمثل عدد الأشخاص الذين يستعملون اليد اليسري، p هو احتمال وجود شخص أعسر، n حجم العينة، فإن X يتبع توزيع ذي الحدين بمعالم $n = 400,\ p = 0.01$

لكن من الملاحظ أن $n > 50$ لذلك فإن X متغير عشوائي سيتبع توزيع بواسون بمعلمة

$$\lambda = np = 400 \times 0.01 = 4$$

$$p(x)=\frac{e^{-\lambda}\lambda^x}{x!}=\frac{e^{-4}4^x}{x!},\qquad x=0,1,2,\dots$$

(أ) احتمال وجود 4 على الأقل يستخدمون اليد اليسرى هو

$$P[X \geq 4] = 1 - P[X < 4]$$

$$= 1 - \sum_{x=0}^{3} p(x) = 1 - \sum_{x=0}^{3} \frac{e^{-4}4^x}{x!}$$

$$= 1 - e^{-4}\left[\frac{4^0}{0!} + \frac{4^1}{1!} + \frac{4^2}{2!} + \frac{4^3}{3!}\right] = 0.5665$$

(ب) العدد المتوقع

$$\mu = E[X] = \lambda = 4$$

مثال (٨-٥)

إذا كان احتمال أن يعاني شخص من أثر سيئ نتيجة الحقن بمحلول معين هو 0.002، فما هو احتمال أن يكون من بين 1000 شخص اختيروا من بين الأشخاص الذين حقنوا بهذا المحلول:

(أ) خمسة فقط يعانون من أثر سيئ.

(ب) خمسة على الأكثر يعانون من أثر سيئ.

(ج) خمسة على الأقل يعانون من أثر سيئ.

الحل

بفرض أن X متغير عشوائي يرمز لعدد الأشخاص الذين يعانون من أثر سيئ نتيجة لعملية الحقن، p احتمال وجود شخص يعاني من أثر سيئ، n حجم العينة المختارة بحيث $n = 1000, \ p = 0.002$

فإن X يتبع توزيع ذي الحدين، لكن $n > 50, p < 0.01$ لذلك فإن X له توزيع بواسون بمعلمة

$$\lambda = np = 1000 \times 0.002 = 2$$

ودالة الكتلة الاحتمالية على الصورة التالية:

$$p(x) = P[X = x] = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-2}2^x}{x!}, \ \ x = 0, 1, 2, \ldots$$

(أ) احتمال أن خمسة أشخاص فقط يعانون من أثر سيئ

$$P[X = 5] = \frac{e^{-2}2^5}{5!} = 0.0361$$

(ب) احتمال أن خمسة أشخاص على الأكثر يعانون من أثر سيئ

$$P[X \le 5] = \sum_{x=0}^{5} p(x) = \sum_{x=0}^{5} \frac{e^{-2}2^x}{x!}$$

$$= e^{-2}\left[\frac{2^0}{0!}+\frac{2^1}{1!}+\frac{2^2}{2!}+\frac{2^3}{3!}+\frac{2^4}{4!}+\frac{2^5}{5!}\right] = 0.9834$$

(ج) احتمال أن خمسة على الأقل يعانون من أثر سيئ

$$P[X \geq 5] = 1 - P[X < 5] = 1 - \sum_{x=0}^{4}\frac{e^{-2}2^x}{x!}$$

$$= 1 - e^{-2}\left[\frac{2^0}{0!}+\frac{2^1}{1!}+\frac{2^2}{2!}+\frac{2^3}{3!}+\frac{2^4}{4!}\right] = 0.05265$$

مثال (٨-٦)

إذا كان متوسط عدد الحوادث الأسبوعية على طريق الرياض مكة هو 4 حوادث, فما هو احتمال وقوع اكثر من حادثتين اسبوعياً؟ ثم أوجد التوقع والانحراف المعياري لعدد الحوادث الأسبوعية على طريق الرياض مكة.

الحل:

بفرض أن X تمثل عدد الحوادث الأسبوعية على طريق الرياض مكة. فإن المتغير العشوائي X يتبع توزيع بواسون بمعلم $\lambda = 4$ حوادث اسبوعياً.

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!},\ \ x = 0, 1, 2, \ldots$$

احتمال وقوع أكثر من حادثتين اسبوعياً هو

$$P[X > 2] = 1 - P[X \leq 2] = 1 - \{P[X = 0] + P[X = 1] + P[X = 2]\}$$
$$= 1 - \left\{\frac{(4)^0 e^{-4}}{0!} + \frac{(4)^1 e^{-4}}{1!} + \frac{(4)^2 e^{-4}}{2!}\right\} = 1 - 0.2381 = 0.7619$$

التوقع والانحراف المعياري لعدد الحوادث الاسبوعية على طريق الرياض مكة

$$\mu = E[X] = \lambda = 4$$

$$\sigma^2 = Var(X) = \lambda = 4$$

٨-٥ التوزيع الهندسي (Geometric Distribution)

إذا كان لدينا محاولة برنوللي وقمنا بتكرار المحاولات عدداً من المرات حتى الحصول على أول نجاح فإنه ينشأ لدينا توزيع جديد يسمى بالتوزيع الهندسي. فإذا عرفنا المتغير العشوائي X على أنه عدد المحاولات التي يجب

إجرائها حتى الحصول على أول نجاح، فإن المتغير العشوائي X يتبع التوزيع الهندسي بمعلمه p وتعطى دالة كتلته الاحتمالية بالصورة الآتية:

$$p(x) = \begin{cases} pq^{x-1}, & x = 1, 2, 3, \ldots \\ 0, & otherwise \end{cases}$$

ومن أمثلة التوزيع الهندسي إذا عرفنا المتغير X بالصورة التالية:

(١) في فحص الإنتاج وكان X عدد السلع التي يتم فحصها حتى الحصول على أول سلعة تالفة.

(٢) عدد مرات الولادة التي تمر بها سيدة حتى الحصول على ذكر.

(٣) في تجربة إلقاء قطعة العملة، إذا عرفنا X بعدد مرات إلقاء قطعة العملة حتى الحصول على أول صورة.

(٤) عدد مرات التصويب على هدف حتى إصابته لأول مرة.

(٥) عدد مرات دخول شخص اختبار القيادة حتى النجاح.

خواص التوزيع الهندسي:

$$1 - \sum_{x} p(x) = 1$$

الإثبات

$$\sum_{x} p(x) = \sum_{x=1}^{\infty} pq^{x-1} = p \sum_{x-1=0}^{\infty} q^{x-1}$$

بوضع $k = x - 1$ فإن

$$\sum_{x} p(x) = p \sum_{k=0}^{\infty} q^{k} = p\left[\frac{1}{1-q}\right] = \frac{p}{p} = 1, \quad |q| < 1$$

حيث

$$\sum_{k=0}^{\infty} a^{k} = 1 + a + a^2 + a^3 + \cdots = \frac{1}{1-a}, \quad |a| < 1$$

وهو مجموع المتتابعة الهندسية إلى مالا نهاية من الحدود.

$$2 - \mu = \frac{1}{p}$$

الإثبات

$$\mu = E[X] = \sum_x xp(x) = \sum_{x=1}^{\infty} xpq^{x-1} = p\sum_{x=1}^{\infty} xq^{x-1}$$

$$= p\sum_{x=0}^{\infty}\frac{d}{dq}\,q^x = p\,\frac{d}{dq}\sum_{x=0}^{\infty} q^x = p\frac{d}{dq}\left[\frac{1}{1-q}\right]$$

$$= p\left[\frac{1}{(1-q)^2}\right] = \frac{1}{p}$$

$$3 - \sigma^2 = \frac{q}{p^2}$$

الإثبات:

$$\because\ \sigma^2 = Var(X) = E[X^2] - \mu^2$$

$$\because\ E[X^2] = E[X(X-1) + X] = E[X(X-1)] + E[X]$$

$$E[X(X-1)] = \sum_x x(x-1)p(x) = \sum_{x=1}^{\infty} x(x-1)pq^{x-1}$$

$$= pq\sum_{x=2}^{\infty} x(x-1)q^{x-2} = pq\sum_{x=0}^{\infty}\frac{d^2}{dq^2}\,q^x = pq\,\frac{d^2}{dq^2}\sum_{x=0}^{\infty} q^x$$

$$= pq\frac{d^2}{dq^2}\left[\frac{1}{1-q}\right] = pq\,\frac{d}{dq}\left[\frac{1}{(1-q)^2}\right]$$

$$= pq\left[\frac{2(1-q)}{(1-q)^4}\right] = \frac{2q}{p^2}$$

$$\Longrightarrow E[X^2] = \frac{2q}{p^2} + \frac{1}{p} = \frac{2q+p}{p^2} = \frac{q+1}{p^2}$$

$$\therefore Var(X) = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}$$

٤- دالة التوزيع للتوزيع الهندسي تعطى بالصورة التالية:

$$F(x) = 1 - q^x, \qquad x = 1, 2, 3, \ldots$$

الإثبات

$$F(x) = P[X \le x] = \sum_{u=1}^{x} p(u)$$

$$= \sum_{u=1}^{\infty} p(u) - \sum_{u=x+1}^{\infty} p(u)$$

$$= \sum_{u=1}^{\infty} p\, q^{u-1} - \sum_{u=x+1}^{\infty} pq^{u-1}$$

$$= p \sum_{u-1=0}^{\infty} q^{u-1} - pq^{x} \sum_{u-x-1=0}^{\infty} q^{u-x-1}$$

$$= p \sum_{r=0}^{\infty} q^{r} - pq^{x} \sum_{k=0}^{\infty} q^{k} = \frac{p}{1-q} - \frac{pq^{x}}{1-q} = 1 - q^{x}, \quad x = 1, 2, \dots$$

إذا كان المتغير X يتبع التوزيع الهندسي بمعلم p فيكتب اختصاراً على الصورة $X \sim Geom(p)$

مثال (٨-٧)

إذا كان احتمال ولادة ذكر في أي ولادة تمر بها سيدة هو 1/3. فما هو احتمال أن

(أ) تضع ذكراً لأول مرة بعد ولادتين.

(ب) تضع ذكراً لأول مرة بعد ثلاث مرات على الأكثر.

الحل

بفرض أن X متغير عشوائي يمثل عدد مرات الولادة التي تمر بها السيدة قبل أن ترزق بأول ذكر.

p احتمال ولادة ذكر حيث $p = \frac{1}{3}$ فإن المتغير X يتبع التوزيع الهندسي وتكون دالة كتلته الاحتمالية على الصورة التالية:

$$p(x) = P[X = x] = pq^{x-1} = \frac{1}{3}\left(\frac{2}{3}\right)^{x-1}, \qquad x = 1, 2, 3, \dots$$

(أ) احتمال أن تضع السيدة ذكراً لأول مرة بعد ولادتين

$$P[X = 2] = \frac{1}{3}\left(\frac{2}{3}\right)^{2-1} = \frac{2}{9}$$

(ب) احتمال أن تضع ذكراً لأول مرة بعد ثلاث مرات على الأكثر

باستخدام دالة الكتلة الاحتمالية

$$P[X \leq 3] = \sum_{x=1}^{3} p(x) = \sum_{x=1}^{3} \frac{1}{3}\left(\frac{2}{3}\right)^{x-1}$$
$$= \frac{1}{3}\left[1 + \frac{2}{3} + \left(\frac{2}{3}\right)^2\right] = \frac{19}{27} = 0.7037$$

باستخدام دالة التوزيع

$$P[X \leq 3] = F(3) = 1 - \left(\frac{2}{3}\right)^3 = 1 - \frac{8}{27} = \frac{19}{27} = 0.7037$$

مثال(٨-٨)

وعاء يحتوي على N من الكرات البيضاء، M من الكرات السوداء، تم اختيار الكرات عشوائياً من الوعاء، كرة في كل مرة حتي يتم الحصول على أول كره سوداء. بفرض أن الكرة المسحوبة سيتم اعادتها للوعاء قبل سحب الكرة التالية. فما هو احتمال

(أ) الحصول على الكرة السوداء في السحبة رقم n

(ب) على الأقل k من العمليات يتم اجرائها قبل الحصول على الكرة السوداء

ثم احسب العدد المتوقع والتباين لمرات السحب من الوعاء للحصول على أول كرة سوداء

الحل:

بفرض أن X يمثل عدد المحاولات التي تتم قبل الحصول أول كرة سوداء، فإن X يتبع التوزيع الهندسي بمعلمة p حيث

$$p = \frac{M}{M+N}, \quad q = 1 - p = \frac{N}{M+N}$$

(أ) احتمال الحصول على الكرة السوداء في السحبة رقم n

$$P[X = n] = \left(\frac{N}{M+N}\right)^{n-1}\left(\frac{M}{M+N}\right) = \frac{MN^{n-1}}{(M+N)^n}$$

(ب) احتمال أنه على الأقل k من العمليات يتم إجراؤها قبل الحصول على الكرة السوداء

$$P[X \geq k] = \sum_{n=k}^{\infty} P[X = n] = \left(\frac{M}{M+N}\right) \sum_{n=k}^{\infty} \left(\frac{N}{M+N}\right)^{n-1}$$

$$= \left(\frac{M}{M+N}\right)\left(\frac{N}{M+N}\right)^{k-1} \Big/ \left[1 - \frac{N}{M+N}\right]$$

$$= \left(\frac{N}{M+N}\right)^{k-1} = \left(1 - \frac{M}{M+N}\right)^{k-1} = (1-p)^{k-1}$$

(ج) العدد المتوقع والتباين لمرات السحب من الوعاء للحصول على أول كرة سوداء

$$\mu = \frac{1}{p} = \frac{M+N}{M}$$

$$\sigma^2 = \frac{q}{p^2} = \left(\frac{N}{M+N}\right) \Big/ \left(\frac{M}{M+N}\right)^2$$

$$= \frac{N(M+N)}{M^2}$$

مثال(٨-٩)

إذا تم إلقاء قطعتي عملة معدنية معاً عدة مرات حتى يتم الحصول على صورة لأول مرة فأوجد

(أ) احتمال الحصول على الصورة رابع محاولة.

(ب) احتمال الحصول على الصورة في رابع محاولة على الأكثر.

(ج) العدد المتوقع للحصول على الصورة لأول مرة

(د) التباين في عدد مرات الرمي حتى الحصول على الصورة

الحل:

بفرض أن A حادثة الحصول على الصورة لأول مرة

$$A = \{(H,T),(T,H)\}$$

وبذلك فإن احتمال الحصول على الصورة عند رمى قطعتي عملة معدنية معاً يساوى

$$p = P(A) = \frac{2}{4} = 0.5$$

بفرض أن X متغير عشوائي يمثل عدد مرات إلقاء قطعتي العملة المعدنية حتى الحصول على الصورة لأول مرة وبالتالي فإن X يتبع التوزيع الهندسي بمعلمة p وتكون دالة الكتلة الاحتمالية له هي:

$$p(x) = pq^{x-1} = (0.5)(0.5)^{x-1} = (0.5)^x, x = 1, 2, 3, \dots$$

(أ) احتمال الحصول على الصورة في رابع محاولة هو

$$P[X = 4] = (0.5)^4 = 0.0625$$

(ب) احتمال الحصول على الصورة في رابع محاولة على الأكثر يساوي

$$P[X \leq 4] = F(4) = 1 - q^4 = 1 - (0.5)^4 = 0.9375$$

(ج) العدد المتوقع للحصول على الصورة لأول مرة يساوي

$$E[X] = \frac{1}{p} = 2$$

(د) التباين في عدد مرات الرمي حتى الحصول على الصورة لأول مرة يساوي

$$Var(X) = \frac{q}{p^2} = \frac{0.5}{(0.5)^2} = 2$$

٨-٦ توزيع ذي الحدين السالب (Negative Binomial Distribution)

يعد توزيع ذي الحدين السالب صورة عامة للتوزيع الهندسي فبدلاً من التركيز على إجراء التجربة حتى الحصول على أول نجاح، فإننا سوف ننظر إلى المتغير العشوائي X على أنه عدد المحاولات اللازمة للحصول على r من النجاح خلال سلسلة من محاولات بيرنوللى. ومن الواضح أنه للحصول على النجاح رقم r يجب أولاً الحصول على $r - 1$ من النجاح في أول $x - 1$ من المحاولات ثم نحصل على النجاح رقم r في المحاولة التالية " رقم x" وعلى ذلك يمكن القول بأن X متغير يتبع توزيعاً يسمى بتوزيع ذي الحدين السالب بمعالم r, p ودالة كتلته الاحتمالية هي:

$$p(x) = \binom{x-1}{r-1} p^{r-1} q^{x-r} p$$

$$= \binom{x-1}{r-1} p^r q^{x-r}, \qquad x = r, r+1, r+2, r+3, \dots$$

حيث إن احتمال الحصول على $r - 1$ نجاح فى $x - 1$ محاولة هو

$$\binom{x-1}{r-1} p^{r-1} q^{x-r}$$

بعض الأمثلة التي يمكن أن يصفها توزيع ذي الحدين السالب:

١- عدد القطع التي يتم فحصها من سلعة معينة حتى يتم الحصول على ثاني قطعة تالفة.

٢- عدد المباريات التي يلعبها فريق ما حتى يفوز بثلاث مباريات.

٣- عدد الطائرات التي تقلع من مطار القاهرة حتى تتأخر طائرتان عن موعد الإقلاع.

خواص توزيع ذي الحدين السالب:

١- $p(x)$ تحقق

$$\sum_{x=-\infty}^{\infty} p(x) = 1$$

الإثبات:

$$\sum_{x=-\infty}^{\infty} p(x) = \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r q^{x-r}$$

$$= p^r \left[1 + \frac{r}{1!} q + \frac{r(r+1)}{2!} q^2 + \frac{r(r+1)(r+2)}{3!} q^3 + \cdots \right]$$

$$= p^r \left[\frac{1}{(1-q)^r}\right] = 1$$

٢- التوقع

$$\mu = \frac{r}{p}$$

الإثبات:

$$\mu = E[X] = \sum_{x=-\infty}^{\infty} x\, p(x) = \sum_{x=r}^{\infty} x \binom{x-1}{r-1} p^r q^{x-r}$$

$$= \sum_{x=r}^{\infty} \frac{x(x-1)!\, p^r q^{x-r}}{(r-1)!\,(x-r)!} = \frac{r}{p} \sum_{x=r}^{\infty} \frac{x!\, p^{r+1}\, q^{x-r}}{r!\,(x-r)!}$$

$$= \frac{r}{p} \sum_{x=r}^{\infty} \binom{x}{r} p^{r+1} q^{x-r} = \frac{r}{p} \sum_{x+1=r+1}^{\infty} \binom{(x+1)-1}{(r+1)-1} p^{r+1} q^{x-r}$$

بوضع $m = x + 1$ وأيضاً $s = r + 1$ فإن

$$= \frac{r}{p} \sum_{m=s}^{\infty} \binom{m-1}{s-1} p^s q^{m-s} = \frac{r}{p} \times 1 = \frac{r}{p}$$

٣- التباين

$$\sigma^2 = \frac{rq}{p^2}$$

الإثبات:

$$\because \sigma^2 = Var(X) = E[X^2] - \mu^2$$

$$\because E[X^2] = E[X(X+1) - X] = E[X(X+1)] - E[X]$$

$$E[X(X+1)] = \sum_{x=-\infty}^{\infty} x(x+1)\, p(x) = \sum_{x=r}^{\infty} x(x+1) \binom{x-1}{r-1} p^r q^{x-r}$$

$$= \sum_{x=r}^{\infty} \frac{x(x+1)(x-1)!\, p^r q^{x-r}}{(r-1)!\,(x-r)!}$$

$$= \sum_{x=r}^{\infty} \frac{r(r+1)(x+1)!\, p^r q^{x-r}}{(r+1)!\,(x-r)!}$$

$$= \frac{r(r+1)}{p^2} \sum_{x+2=r+2}^{\infty} \binom{(x+2)-1}{(r+2)-1} p^{r+2} q^{x-r}$$

بوضع $y = x + 2$ وأيضاً $k = r + 2$ فإن

$$E[X(X+1)] = \frac{r(r+1)}{p^2} \sum_{y=k}^{\infty} \binom{y-1}{k-1} p^k q^{y-k} = \frac{r(r+1)}{p^2}$$

$$\therefore E[X^2] = \frac{r(r+1)}{p^2} - \frac{r}{p} = \frac{r}{p^2}(r + 1 - p) = \frac{r(r+q)}{p^2}$$

$$\therefore \sigma^2 = \frac{r^2 + rq}{p^2} - \left(\frac{r}{p}\right)^2 = \frac{r^2 + rq - r^2}{p^2} = \frac{rq}{p^2}$$

من الواضح أنه في توزيع ذي الحدين السالب إذا كانت $r = 1$ فإننا سوف نحصل على التوزيع الهندسي.

مثال(٨–١٠)

إذا كان احتمال إصابة هدف ما بنجاح هو 0.6. وبفرض أن هناك عدة محاولات للتصويب على الهدف حتى الحصول على ثلاثة عمليات إصابة بنجاح. أوجد

(أ) احتمال أن يكون هناك 6 محاولات ضرورية.

(ب) احتمال أن يكون هناك أقل من 6 محاولات.

(ج) العدد المتوقع لعمليات الإطلاق.

الحل:

بفرض أن X متغير عشوائي يمثل عدد المحاولات اللازمة للحصول على 3 عمليات إصابة ناجحة للهدف وبالتالي فإن X متغير عشوائي يتبع توزيع ذي الحدين السالب بمعالم $r = 3, p = 0.6$ ودالة كتلته الاحتمالية هي:

$$p(x) = \binom{x-1}{2}(0.6)^3(0.4)^{x-3}, \qquad x = 3,4,5,\ldots$$

(أ) احتمال أن يكون هناك 6 محاولات ضرورية هو

$$P[X = 6] = \binom{6-1}{2}(0.6)^3(0.4)^{6-3} = 0.1382$$

(ب) احتمال أن يكون هناك أقل من 6 محاولات هو

$$\begin{aligned} P[X < 6] &= p(3) + p(4) + p(5) \\ &= \binom{2}{2}(0.6)^3(0.4)^0 + \binom{3}{2}(0.6)^3(0.4)^1 + \binom{4}{2}(0.6)^3(0.4)^2 \\ &= 0.6826 \end{aligned}$$

(ج) العدد المتوقع لعمليات الإطلاق يساوي

$$E[X] = \frac{r}{p} = \frac{3}{0.6} = 5$$

٨–٧ التوزيع الهندسي الزائدي (Hyper Geometric Distribution)

إذا كان لدينا مجتمع يحتوى على N من العناصر منها M من العناصر تنتمي للفئة C والباقي $N - M$ من العناصر تنتمي لمكملة الفئة C. سحبت من هذا المجتمع عينة عشوائية حجمها n، وتم تعريف المتغير العشوائي X

بحيث يمثل عدد عناصر العينة التي تنتمي للفئة C فإن المتغير X يتبع التوزيع الهندسي الزائدي والتي تعطى دالة كتلته الاحتمالية بالصورة التالية:

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, x = 0, 1, \dots, n$$

ونلاحظ أن نسبة العناصر التي تنتمي إلى الفئة C هي

$$p = \frac{M}{N} \Rightarrow M = Np$$

كذلك نسبة العناصر التي تنتمي لمكملة الفئة C هي

$$q = \frac{N-M}{N} \Rightarrow N - M = Nq$$

وبالتالي فإن دالة الكتلة الاحتمالية يمكن كتابتها على الصورة التالية:

$$p(x) = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}, x = 0, 1, \dots, n$$

خواص التوزيع الهندسي الزائدي:

١- $p(x)$ تحقق العلاقة

$$\sum_{x=0}^{n} p(x) = 1$$

الإثبات:

$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{x=0}^{n} \binom{Np}{x}\binom{Nq}{n-x}$$

وباستخدام المتطابقة

$$\binom{a+b}{n} = \sum_{j=0}^{n} \binom{a}{j}\binom{b}{n-j}$$

حيث a, b أعداد صحيحة موجبة, $n < a + b$ نجد أن:

$$\sum_{x=0}^{n} p(x) = \frac{1}{\binom{N}{n}}\binom{Np+Nq}{n} = \frac{1}{\binom{N}{n}}\binom{N}{n} = 1$$

٢- التوقع الرياضي

$$\mu = np$$

الإثبات:

$$\mu = E[X] = \sum_{x=0}^{n} xp(x) = \frac{1}{\binom{N}{n}}\sum_{x=0}^{n} x\binom{Np}{x}\binom{Nq}{n-x}$$

$$= \frac{Np}{\binom{N}{n}}\sum_{x=1}^{n}\frac{(Np-1)!}{(x-1)!\,(Np-x)!}\ \binom{Nq}{n-x}$$

$$= \frac{Np}{\binom{N}{n}}\sum_{x-1=0}^{n}\binom{Np-1}{x-1}\binom{Nq}{n-1-(x-1)}$$

بوضع $y = x - 1$ فإن

$$\mu = \frac{Np}{\binom{N}{n}}\sum_{y=0}^{n-1}\binom{Np-1}{y}\binom{Nq}{n-1-y}$$

$$= \frac{Np}{\binom{N}{n}}\binom{N-1}{n-1} = \frac{Np(N-1)!\,n!\,(N-n)!}{N!\,(n-1)!\,(N-n)!} = np$$

٣- التباين

$$\sigma^2 = \left(\frac{N-n}{N-1}\right)npq$$

الإثبات:

$$\because \sigma^2 = Var(X) = E[X^2] - \mu^2$$

$$\because E[X^2] = E[X(X-1)] + E[X]$$

$$E[X(X-1)] = \sum_{x=1}^{n} x(x-1)p(x)$$

$$=\frac{1}{\binom{N}{n}}\sum_{x=1}^{n}x(x-1)\binom{Np}{x}\binom{Nq}{n-x}$$

$$=\frac{Np(Np-1)}{\binom{N}{n}}\sum_{x=2}^{n}\frac{(Np-2)!}{(x-2)!\,(Np-x)!}\binom{Nq}{n-x}$$

$$=\frac{Np(Np-1)}{\binom{N}{n}}\sum_{x-2=0}^{n}\binom{Np-2}{x-2}\binom{Nq}{n-2-(x-2)}$$

بوضع $y=x-2$ فإن

$$E[X(X-1)]=\frac{Np(Np-1)}{\binom{N}{n}}\sum_{y=0}^{n-2}\binom{Np-2}{y}\binom{Nq}{n-2-y}$$

$$=\frac{Np(Np-1)}{\binom{N}{n}}\binom{N-2}{n-2}$$

$$=\frac{n(n-1)p(Np-1)}{N-1}$$

$$\Rightarrow E[X^2]=\frac{n(n-1)p(Np-1)}{N-1}+np$$

$$\Rightarrow \sigma^2=E[X^2]-\mu^2=\frac{n(n-1)p(Np-1)}{N-1}+np-n^2p^2$$

$$=\frac{np}{N-1}[nNp-n-Np+1+N-1-NnP+nP]$$

$$=\left(\frac{N-n}{N-1}\right)npq$$

مثال (٨-١١)

مستودع لشركة أجهزة حاسب آلي به 100 جهاز حاسب آلي منها 20 من النوع توشيبا . سحبت منه عشوائياً عينة من 10 أجهزة. أوجد احتمال أن تكون العينة

(أ) تحتوي على جهاز واحد من النوع توشيبا على الأكثر.

(ب) لا تحتوي على أجهزة من النوع توشيبا.

(ج) العدد المتوقع من الأجهزة من النوع توشيبا في العينة.

الحل

بفرض أن X تمثل عدد اجهزة الحاسب من النوع توشيبا في العينة، فإن X له التوزيع الهندسي الزائدي وبالتالي فإن دالة الكتلة الاحتمالية هي:

$$p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, \ldots, n$$

حيث

$$N = 100, \qquad M = 20, \qquad n = 10$$

$$\Rightarrow p(x) = \frac{\binom{20}{x}\binom{80}{10-x}}{\binom{100}{10}}, \qquad x = 0, 1, \ldots, 10$$

(أ) احتمال أن العينة تحتوى على جهاز واحد من النوع توشيبا على الأكثر يساوي

$$P[X \leq 1] = P[X = 0] + P[X = 1]$$

$$= \frac{\binom{20}{0}\binom{80}{10}}{\binom{100}{10}} + \frac{\binom{20}{1}\binom{80}{9}}{\binom{100}{10}} = 0.0951 + 0.2679 = 0.3630$$

(ب) احتمال أن العينة لا تحتوى على أجهزة من النوع توشيبا هو

$$P[X = 0] = \frac{\binom{20}{0}\binom{80}{10}}{\binom{100}{10}} = 0.0951$$

(ج) العدد المتوقع من الأجهزة من النوع توشيبا في العينة

$$\mu = np = \frac{nM}{N} = \frac{10(20)}{100} = 2$$

أسئلة وتمارين (٨)

١- ما المقصود بالتوزيعات الاحتمالية المنفصلة؟

٢- متى تنشأ تجربة ذات الحدين؟

٣- ما هي العلاقة بين كلّ من

(أ) توزيع ذات الحدين وتوزيع برنوللى.

(ب) التوزيع الهندسي وتوزيع ذي الحدين السالب.

٤- أوجد قيمة التوقع الرياضي والتباين لتوزيع ذي الحدين السالب؟ ثم منه عين التوقع الرياضي والتباين للتوزيع الهندسي؟

٥- إذا كان احتمال أن يرصد جهاز رادار طائرة معاديه هو 0.9 وكانت لدينا خمسة أجهزة من هذا النوع تعمل مستقلة عن بعضها البعض فأوجد قيم احتمالات الحوادث التالية:

(أ) أربعة أجهزة ترصد الطائرة.

(ب) على الأقل جهاز واحد يرصد الطائرة.

٦- اثبت أن توزيع ذي الحدين بمعالم n, p يقترب من توزيع بواسون عندما $n \longrightarrow \infty,\ p \longrightarrow 0$

٧- احتمال أن يكسب الفريق A في أي مباراة يلعبها هو 2/3 لعب الفريق A أربع مباريات. أوجد احتمال أن يفوز في

(أ) مباريتين فقط.

(ب) على الأقل مباراة واحدة.

(ج) أكثر من نصف المباريات.

٨- إذا كان المتغير العشوائي X يرمز لعدد مرات إصابة الهدف خلال 8 محاولات لشخص ما. وكان احتمال إصابة الهدف في كل محاولة هو 1/4 أوجد احتمال

(أ) إصابة الهدف مرتين على الأقل.

(ب) إصابة الهدف مرة واحدة على الأكثر.

(ج) التوقع الرياضي والتباين لعدد مرات إصابة الهدف.

٩- إذا كان معلوماً أن احتمال المعاناة من أثار جانبية بعد استخدام أحد الأمصال هو 0.001 وتم تطعيم 1000 شخص بذلك المصل. أوجد احتمال أن

(أ) واحد على الأكثر يعاني من الآثار الجانبية.

(ب) أربع أو خمس أو ستة أشخاص يعانون من الآثار الجانبية.

١٠- أوجد $P[1 \leq X \leq 4]$ للمتغير العشوائي X إذا كان X يتبع توزيع:

(أ) بواسون بمعلمة $\lambda=1$

(ب) توزيع ذو الحدين بمعالم $n = 10,\ p = 0.1$

(ج) التوزيع الهندسي بمعلمة $p = \frac{1}{3}$

(د) توزيع ذي الحدين السالب بمعالم $r = 3, p = \frac{1}{3}$

١١- ألقي زهر نرد متزن خمس مرات متتالية. فإذا كان X متغير عشوائي يمثل عدد مرات الحصول على رقم أقل من 3 أوجد احتمال أن يأخذ هذا المتغير القيمة 4

١٢- إذا كان X متغيراً عشوائياً يتبع التوزيع الثنائي، حيث $p = 0.25$ فعين أقل قيمة لـ n إذا كان $P[X \geq 1] = 0.70$

١٣- إذا كان هناك 300 خطأً مطبعياً موزعة عشوائياً على كتاب به 500 صفحة. أوجد احتمال أن يكون بإحدى الصفحات

(أ) خطآن مطبعيان.

(ب) خطآن مطبعيان على الأكثر.

١٤- إذا كان X متغيراً عشوائياً يتبع توزيع بواسون بحيث $P[X = 0] = P[X = 1]$ فأوجد $P[X = 2]$

١٥- إذا كان احتمال إصابة شخص ما لهدف هو 0.3 وأتيحت له فرصة التصويب 8 مرات فأوجد

(أ) احتمال إصابة الهدف ثلاثة مرات على الأكثر.

(ب) احتمال إصابة الهدف مرتين؟

(ج) احتمال إصابة الهدف مرتين على الأقل؟

(د) ما هو العدد المتوقع لإصابته للهدف؟

١٦- إذا كان احتمال أن ينجح طالب في مقرر 1080 أحص هو 0.825 فما هو احتمال نجاح 85% من الطلاب في شعبة تحتوى على 40 طالب؟

١٧- إذا كان عدد الحوادث الأسبوعية على طريق الرياض – الخرج هو ثمانية حوادث فأوجد

(أ) احتمال وقوع حادثتين في يوم ما من الأسبوع.

(ب) احتمال وقوع أربعة حوادث على الأكثر خلال يومين من الأسبوع.

(ج) العدد المتوقع للحوادث خلال 12 أسبوع على طريق الرياض – الخرج؟

١٨- في إحدى رحلات الطيران والتي تحتوى على 400 راكب فقد وجد شخص مصاب بمرض أنفلونزا الخنازير، فإذا كان احتمال انتقال الفيروس من شخص مصاب لشخص سليم هو 0.0005 فعند وصول الرحلة إلى المكان المحدد لها فما هو احتمال

(أ) وجود ثلاثة أشخاص مصابين بالمرض.

(ب) وجود شخصين على الأكثر مصابين بالمرض.

(ج) عدم وجود أشخاص مصابه بالمرض.

(د) العدد المتوقع للإصابة بالمرض.

١٩- يلعب شخصان القمار بواسطة زهرة النرد وقد اتفقا على أن يرمى أحدهما الزهرة فإذا حصل على 3 أو 6 كسب عشرة دولارات. أما إذا حصل على أي عدد أخر فإنه يخسر خمسة دولارات. فإذا استمر اللعب 30 دوراً فاحسب الاحتمالات الآتية:

(أ) أن يكسب الرامي 75 دولاراً.

(ب) أن يخسر الرامي 75 دولاراً.

(ج) أن لا يخسر الرامي أو يكسب.

(د) أن يخسر الرامي.

٢٠- أسرة بها خمسة أطفال فإذا كان X متغير عشوائي يمثل عدد الأولاد الذكور في هذه الأسرة. فأوجد احتمال

(أ) أن يكون عدد الأولاد الذكور أقل من واحد.

(ب) أن يتراوح عدد الأولاد الذكور بين 2، 4

(ج) أن يكون عدد الأولاد الذكور أكثر من ثلاثة.

٢١- صممت قطعة عملة بحيث إن احتمال الحصول على صورة يساوى $\frac{2}{3}$ فإذا ألقيت قطعة العملة عدة مرات حتى الحصول على الصورة ثلاثة مرات فما هو احتمال

(أ) أن نحصل على 3 صور بعد الرمية الخامسة على الأكثر.

(ب) الحصول على 3 صور بعد الرمية العاشرة على الأقل.

(ج) العدد المتوقع لمرات إلقاء قطعة العملة وأيضا التباين.

٢٢- إذا كان لدينا شعبة في كلية العلوم والدراسات الإنسانية بالخرج تحتوى على 10 طلاب يقيمون في الرياض، 25 طالباً يقيمون في محافظة الخرج، فإذا اخترنا عينة مكونة من 5 طلاب من هذه الشعبة لتمثيل طلاب الشعبة في الملتقى الأسبوعي بعميد الكلية، فما هو احتمال أن

(أ) تحتوى العينة على طالبين من طلاب الرياض.

(ب) لا تحتوى العينة على طلاب من الخرج.

(ج) تحتوى العينة على ثلاثة طلاب من الرياض على الأكثر.

(د) ما هو العدد المتوقع من الطلاب المقيمين بالرياض بالعينة.

٢٣- في اختبار متعدد الإجابات من 6 أسئلة وضع لكل منها 3 إجابات واحدة منها فقط صحيحة، اتخذ طالب السياسة التالية:

يرمى زهرة النرد المتزنة فإذا كان الرقم الظاهر 1 أو 2 قرر اختيار الإجابة الأولى. وإذا كانت نتيجة الزهرة 3 أو 4 قرر اختيار الإجابة الثانية وخلاف ذلك يقرر اختيار الإجابة الثالثة. احسب احتمال

(أ) الحصول على ثلاثة إجابات صحيحة فقط.

(ب) عدم الحصول على إجابة صحيحة.

(ج) ما هو العدد المتوقع للإجابات الصحيحة التي سيحصل عليها الطالب باتباعه هذه السياسة.

٢٤- أعلنت وزارة الصحة عن توفر 4 بعثات لدراسة طب الأسنان فتقدم لها 10 رجال، 6 من النساء، وعند الاختيار وجد أنهم جميعاً متساوون في المؤهل والخبرة فتقرر إتباع طريقة الاختيار العشوائي. فأوجد

(أ) التوزيع الاحتمالي لعدد الرجال المختارين.

(ب) احتمال اختيار رجل واحد على الأقل.

(ج) احتمال اختيار رجل واحد على الأكثر.

٢٥- بفرض أن احتمال أن تنتج آلة معينة في مصنع عنصراً معيباً هو 0.1 فأوجد احتمال أن تحتوى عينة مكونة من 10 عناصر من إنتاج تلك الآلة على عنصرين على الأكثر معيبين؟

٢٦- بفرض أن لدينا تجربة فيزيائية تستخدم لتعيين عدد أشعة ألفا في الثانية لكتلة واحد جرام من مادة مشعة. فإذا كان معلوماً من التجارب السابقة أنه في المتوسط 3.2 من أشعة ألفا تنبعث من تلك المادة في الثانية الواحدة فما هو احتمال أن لا يعد الجهاز أكثر من شعاعين من أشعة ألفا في الثانية الواحدة؟

٢٧- إذا كان X متغيراً عشوائياً يتبع توزيع بواسون بمعلم λ فاثبت أن

$$E[X^n] = \lambda\, E[(X+1)^{n-1}]$$

ثم استخدم النتيجة السابقة لتعيين قيمة $E[X^3]$

٢٨- إذا كان X متغيراً عشوائياً يتبع توزيع ذي الحدين بمتوسط 6 وتباين 2.4 فأوجد قيمة $P[X = 5]$

٢٩- إذا كان العدد المتوقع من الأخطاء المطبعية في صفحة لمجلة ما هو 0.2 فما هو احتمال أن تكون الصفحة التالية:

(أ) لا تحتوى على أخطاء مطبعية.

(ب) تحتوى على خطأين على الأكثر.

٣٠- ما هو العدد المتوقع والتباين لمرات إلقاء زهرة النرد للحصول على الرقم واحد أربعة مرات؟

الفصل التاسع

بعض التوزيعات الاحتمالية المتصلة

SOME CONTINUOUS PROBABILITY DISTRIBUTIONS

٩-١ مقدمة

تعرضنا في الفصل السابق لبعض التوزيعات الاحتمالية للمتغيرات المنفصلة والتي لها قيم معدودة سواء كانت قيم محدودة أو غير محدودة والتي تسمى بالتوزيعات الاحتمالية المنفصلة، ولكن إذا كان المتغير يأخذ قيماً غير معدودة وغير نهائية تقع داخل فترة ما فيسمى (متغير عشوائي متصل). لذا سوف نستعرض في هذا الفصل بعض التوزيعات الاحتمالية للمتغيرات المتصلة وسنهتم بالتوزيعات ذات التطبيقات المهمة وسنقوم بعرض بعض خواص تلك التوزيعات والتوقع الرياضي والتباين لتلك التوزيعات.

٩-٢ التوزيع المنتظم (Uniform Distribution)

يعتبر التوزيع المنتظم المتصل من أبسط التوزيعات الإحصائية، وتتميز دالة كثافته الاحتمالية بأنها منتظمة (ثابتة) على فترة مغلقة (a, b). فيقال إن المتغير العشوائي X يتبع التوزيع المنتظم على الفترة (a, b) إذا كانت دالة كثافة احتماله تعطى بمقدار ثابت في الفترة (a, b) وتعرف كالآتي:

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

واضح أن $f(x) \ge 0$ لجميع قيم x

خواص التوزيع المنتظم:

$1-\int_{-\infty}^{\infty} f(x)dx = 1$

الإثبات

$$\int_{-\infty}^{\infty} f(x)dx = \int_{a}^{b} \frac{1}{b-a}dx = \left[\frac{x}{b-a}\right]_a^b = \frac{b-a}{b-a} = 1$$

$2-\mu = \frac{b+a}{2}$

الإثبات

$$\mu = E[X] = \int_{-\infty}^{\infty} x\, f(x)dx = \int_{a}^{b} \frac{x}{b-a}dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b$$

$$= \frac{1}{2(b-a)}(b^2-a^2) = \frac{b+a}{2}$$

$3-\sigma^2 = \frac{(b-a)^2}{12}$

الإثبات

$$\because \sigma^2 = Var(X) = E[X^2] - \mu^2$$

$$\Longrightarrow E[X^2] = \int_{-\infty}^{\infty} x^2\, f(x)dx = \int_{a}^{b} \frac{x^2}{b-a}dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b$$

$$= \frac{1}{3(b-a)}(b^3-a^3) = \frac{b^2+ab+a^2}{3}$$

$$\therefore \sigma^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2$$

$$= \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4}$$

$$= \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}$$

٤- دالة التوزيع للمتغير العشوائي X الذي له التوزيع المنتظم هي

$$F(x)=\begin{cases}0, & x<a\\ \dfrac{x-a}{b-a}, & a\le x\le b\\ 1, & x\ge b\end{cases}$$

الإثبات

إذا كانت $x < a$ فإن

$$F(x)=P[X\le x]=\int_{-\infty}^{x} f(u)du=\int_{-\infty}^{x} 0\,du=0$$

إذا كانت $a \le x \le b$ فإن

$$F(x)=P[X\le x]=\int_{-\infty}^{x} f(u)du$$

$$=\int_{-\infty}^{a} f(u)du+\int_{a}^{x} f(u)du$$

$$=\int_{-\infty}^{a} 0\,du+\int_{a}^{x}\frac{1}{b-a}\,du=\frac{x-a}{b-a}$$

إذا كانت $x \ge b$ فإن

$$F(x)=P[X\le x]=\int_{-\infty}^{x} f(u)du$$

$$=\int_{-\infty}^{a} f(u)du+\int_{a}^{b} f(u)du+\int_{b}^{x} f(u)du$$

$$=\int_{-\infty}^{a} 0\,du+\int_{a}^{b} f(u)du+\int_{b}^{x} 0\,du=1$$

$$\Rightarrow F(x)=\begin{cases}0, & x<a\\ \dfrac{x-a}{b-a}, & a\le x\le b\\ 1, & x\ge b\end{cases}$$

مثال (٩-١)

إذا كان X متغيراً عشوائياً يتبع التوزيع المنتظم على الفترة $(1,3)$. فأوجد

$$P[1 \le X \le 2.6], P[X \ge 2]$$

الحل

بما أن X له التوزيع المنتظم على الفترة $(1,3)$ فإن دالة الكثافة للمتغير X تأخذ الشكل التالي:

$$f(x) = \begin{cases} \frac{1}{2}, & 1 \le x \le 3 \\ 0, & otherwise \end{cases}$$

ودالة التوزيع

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{x-1}{2}, & 1 \le x \le 3 \\ 1, & x \ge 3 \end{cases}$$

وبالتالي فإن الاحتمالات المطلوبة يمكن تعيينها باستخدام دالة التوزيع كما يلي:

$$P[1 \le X \le 2.6] = F(2.6) - F(1) = 0.8$$

$$P[X \ge 2] = 1 - P[X < 2] = 1 - F(2) = 0.5$$

ويمكن حساب الاحتمالات السابقة باستخدام دالة الكثافة الاحتمالية كالتالي:

$$P[1 \le X \le 2.6] = \int_1^{2.6} f(x)dx = \int_1^{2.6} \frac{1}{2}\, dx = \left[\frac{x}{2}\right]_1^{2.6} = 0.8$$

$$P[X \ge 2] = \int_2^{3} f(x)dx = \int_2^{3} \frac{1}{2}\, dx = \left[\frac{x}{2}\right]_2^{3} = 0.5$$

مثال (٩-٢)

إذا كان المتغير العشوائي X يتبع التوزيع المنتظم على الفترة $(0, 10)$.

(أ) أوجد قيم الاحتمالات $P[X < 3], P[X > 6], P[3 < X < 8]$

(ب) ارسم دالة التوزيع $F(x)$

(ج) أوجد التوقع الرياضي والتباين للمتغير X

الحل

دالة الكثافة الاحتمالية للمتغير X على الصورة التالية

$$f(x) = \begin{cases} \frac{1}{10}, & 0 \leq x \leq 10 \\ 0, & otherwise \end{cases}$$

(أ) يمكن حساب قيمة الاحتمالات كالتالي:

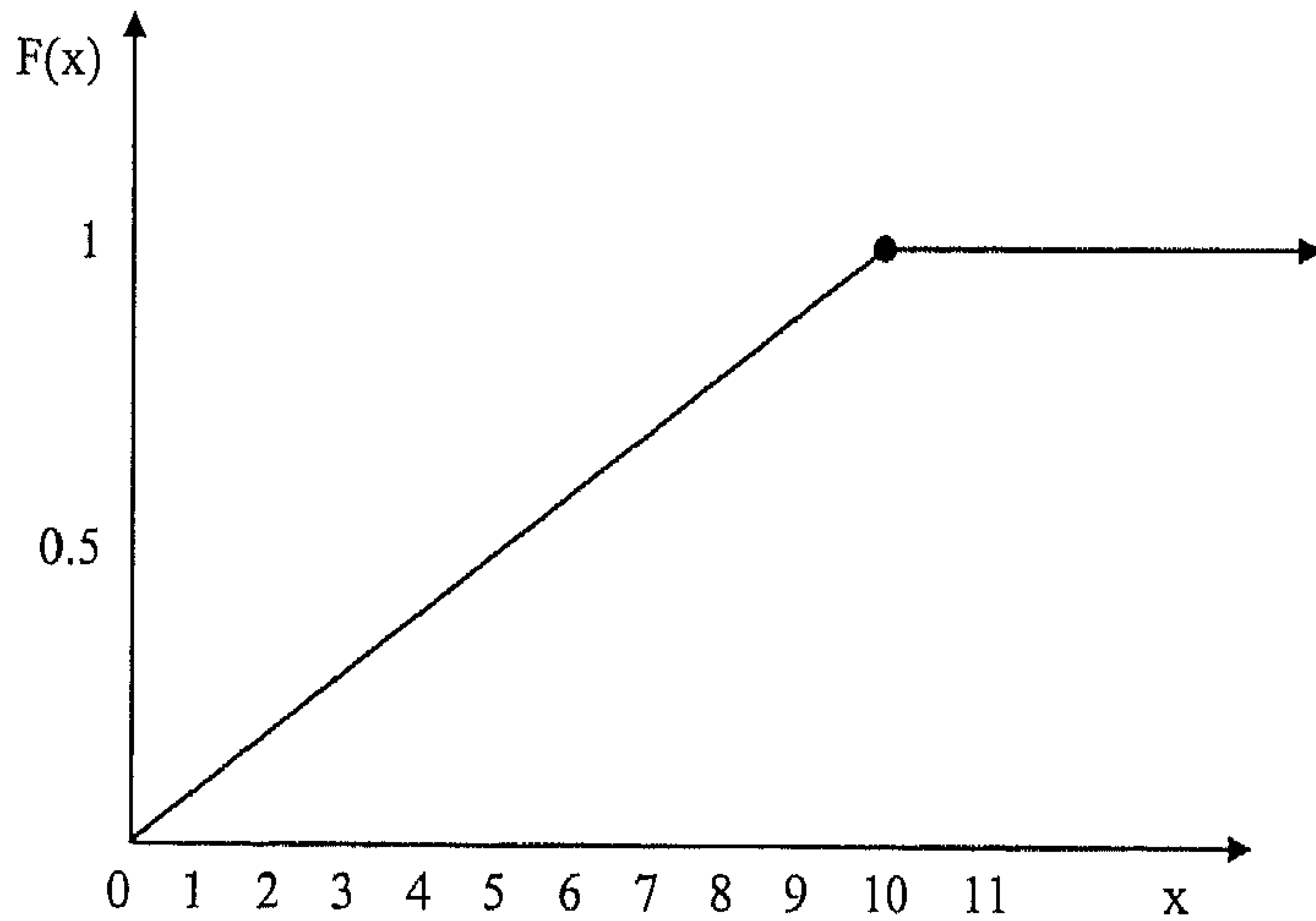
$$P[X < 3] = \int_{-\infty}^{3} f(x)\, dx = \int_{0}^{3} \frac{dx}{10} = \left[\frac{x}{10}\right]_0^3 = \frac{3}{10} = 0.3$$

$$P[X > 6] = \int_{6}^{\infty} f(x)\, dx = \int_{6}^{10} \frac{1}{10}\, dx = \left[\frac{x}{10}\right]_6^{10} = \frac{4}{10} = 0.4$$

$$P[3 < X < 8] = \int_{3}^{8} f(x)\, dx = \int_{3}^{8} \frac{dx}{10} = \left[\frac{x}{10}\right]_3^{8} = \frac{5}{10} = 0.5$$

(ب) دالة التوزيع تعطي بالصورة التالية:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{10}, & 0 \leq x \leq 10 \\ 1, & x \geq 10 \end{cases}$$



(ج) التوقع الرياضي والتباين:

$$E[X] = \frac{a+b}{2} = \frac{10+0}{2} = 5$$

$$Var(X) = \frac{(b-a)^2}{12} = \frac{(10-0)^2}{12} = \frac{100}{12} = 8.333$$

٩-٣ التوزيع الأسي (Exponential Distribution)

التوزيع الأسي له تطبيقات واسعة في مجال الإحصاء وعلى الأخص في نظرية الموثوقية (Reliability Theory) فيصف هذا التوزيع كثيراً من الظواهر مثل أعمار بعض السلع الكهربائية، الوقت اللازم حتى تتعطل بعض الأنظمة الإلكترونية، وقت الانتظار اللازم لوقوع حدث معين.

ويقال للمتغير العشوائي X إنه يتبع التوزيع الأسي بمعلمة λ حيث $0 < \lambda$ إذا كانت دالة كثافة احتماله تعطى بالصورة الآتية:

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

خواص التوزيع الأسي:

$$1- \int_{-\infty}^{\infty} f(x)dx = 1$$

الإثبات

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} \lambda\, e^{-\lambda x}dx = \lambda\left[-\frac{e^{-\lambda x}}{\lambda}\right]_0^{\infty} = -[e^{-\infty} - e^0] = 1$$

$$2- \mu = \frac{1}{\lambda}$$

الإثبات

$$\mu = E[X] = \int_{-\infty}^{\infty} x\, f(x)dx = \int_{0}^{\infty} x\lambda\, e^{-\lambda x}\, dx$$

بالتكامل بالتجزيء حيث

$$\int u\, dv = uv - \int v\, du$$

وبفرض

$$u = x, \quad dv = \lambda e^{-\lambda x} dx \Rightarrow du = dx, v = -e^{-\lambda x}$$

$$\therefore E[X] = -\left[x\, e^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x}\, dx = 0 + \int_0^\infty e^{-\lambda x}\, dx$$

$$= -\frac{1}{\lambda}\left[e^{-\lambda x}\right]_0^\infty = -\frac{1}{\lambda}\left[e^{-\infty} - e^0\right] = \frac{1}{\lambda}$$

$$3 - \sigma^2 = \frac{1}{\lambda^2}$$

الإثبات

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^\infty x^2 \lambda\, e^{-\lambda x}\, dx$$

بالتكامل بالتجزيء حيث

$$u = x^2, \quad dv = \lambda e^{-\lambda x} dx \Rightarrow du = 2x dx, v = -e^{-\lambda x}$$

$$\therefore E[X^2] = -2\left[x^2\, e^{-\lambda x}\right]_0^\infty + 2\int_0^\infty x\, e^{-\lambda x}\, dx = 0 + 2\int_0^\infty x e^{-\lambda x}\, dx$$

$$= 2\int_0^\infty x e^{-\lambda x}\, dx$$

بالتكامل بالتجزيء مرة ثانية حيث

$$u = x, \quad dv = e^{-\lambda x} dx \Rightarrow du = dx, v = -\frac{1}{\lambda} e^{-\lambda x}$$

$$\therefore E[X^2] = -\frac{2}{\lambda}\left[x\, e^{-\lambda x}\right]_0^\infty + \frac{2}{\lambda}\int_0^\infty e^{-\lambda x}\, dx = 0 + \frac{2}{\lambda}\int_0^\infty e^{-\lambda x}\, dx$$

$$= -\frac{2}{\lambda^2}\left[e^{-\lambda x}\right]_0^\infty = -\frac{2}{\lambda^2}\left[e^{-\infty} - e^0\right] = \frac{2}{\lambda^2}$$

$$\therefore \sigma^2 = E[X^2] - \mu^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

٤- دالة التوزيع للمتغير العشوائي X هي

$$F(x)=\begin{cases}0, & x<0\\ 1-e^{-\lambda x}, & x\geq 0\end{cases}$$

الاثبات

- اذا كانت $x<0$ فان

$$F(x)=P[X\leq x]=\int_{-\infty}^{x}f(u)du=\int_{0}^{x}0\,du=0$$

- اذا كانت $x\geq 0$ فان

$$F(x)=P[X\leq x]=\int_{-\infty}^{x}f(u)du=\int_{0}^{x}\lambda e^{-\lambda u}du=-\left[e^{-\lambda u}\right]_{0}^{x}=1-e^{-\lambda x}$$

فإن دالة التوزيع هي

$$F(x)=\begin{cases}0, & x<0\\ 1-e^{-\lambda x}, & x\geq 0\end{cases}$$

مثال (٩-٣)

إذا كان X متغيراً عشوائياً دالة كثافته الاحتمالية هي

$$f(x)=\begin{cases}ce^{-2x}, & x\geq 0\\ 0, & x<0\end{cases}$$

فأوجد

(أ) قيمة الثابت c

(ب) قيمة $P[X>2]$

الحل

(أ) حيث إن $f(x)$ دالة كثافة احتمالية فإن

$$\int_{-\infty}^{\infty}f(x)dx=1$$

$$\int_{-\infty}^{\infty}f(x)dx=\int_{0}^{\infty}ce^{-2x}dx=-\frac{c}{2}[e^{-2x}]_{0}^{\infty}$$

$$=-\frac{c}{2}[e^{-\infty}-e^{0}]=\frac{c}{2}=1$$

$$\Longrightarrow c = 2$$

$$\therefore f(x) = \begin{cases} 2e^{-2x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

(ب) قيمة الاحتمال $P[X > 2]$

$$P[X > 2] = \int_{2}^{\infty} f(x)dx = \int_{2}^{\infty} 2e^{-2x}dx = -[e^{-2x}]_{2}^{\infty} = e^{-4} = 0.0183$$

مثال (٩-٤)

بفرض أن طول المكالمة الهاتفية بالدقائق هو متغير عشوائي X يتبع التوزيع الأسي

$$f(x) = \begin{cases} c\, e^{-\frac{x}{10}}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

فإذا وجد شخص ما يقوم بإجراء اتصال قبل وصولك مما يتسبب في الانتظار حتي ينتهي. فأوجد

(أ) قيمة الثابت c

(ب) احتمال، انك سوف تنتظر: اكثر من 10 دقائق، بين 10، 20 دقيقة

(ج) التوقع والتباين لوقت الانتظار.

الحل

(أ) حيث ان $f(x)$ دالة كثافة احتمالية فان

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} ce^{-\frac{x}{10}}dx = -10c\left[e^{-\frac{x}{10}}\right]_{0}^{\infty}$$

$$= -10c[e^{-\infty} - e^{0}] = 10c = 1$$

$$\therefore c = \frac{1}{10}$$

ومما سبق فإن

$$\therefore f(x) = \begin{cases} \frac{1}{10} e^{-\frac{x}{10}}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

(ب) احتمال الانتظار أكثر من 10 دقائق

$$P[X > 10] = \int_{10}^{\infty} f(x)dx$$

$$= \int_{10}^{\infty} \frac{1}{10} e^{-\frac{x}{10}} dx = \left[-e^{-\frac{x}{10}} \right]_{10}^{\infty}$$

$$= -(e^{-\infty} - e^{-1}) = e^{-1} = 0.3679$$

(ب) احتمال الانتظار بين 10، 20 دقيقة

$$P[10 < X < 20] = \int_{10}^{20} f(x)dx$$

$$= \int_{10}^{20} \frac{1}{10} e^{-\frac{x}{10}} dx = \left[-e^{-\frac{x}{10}} \right]_{10}^{20}$$

$$= -(e^{-2} - e^{-1}) = e^{-1} - e^{-2}$$

$$= 0.3679 - 0.1353 = 0.2326$$

(ج) حيث ان X يتبع التوزيع الأسي فإن التوقع والتباين لوقت الانتظار.

$$\therefore \mu = E[X] = \frac{1}{\lambda} = \frac{1}{\frac{1}{10}} = 10$$

$$\therefore \sigma^2 = \frac{1}{\lambda^2} = \frac{1}{(0.1)^2} = 100$$

٩-٤ التوزيع الطبيعي (Normal Distribution)

التوزيع الطبيعي من أفضل وأكثر التوزيعات الاحتمالية استخداماً؛ فهو يستخدم بكثرة في الأبحاث المالية والعلمية. وفى الواقع فهو يؤدي دوراً أعم وأشمل من ذلك بكثير، فهو علم بارز بين التوزيعات الاحتمالية بمختلف أنواعها ومسمياتها وتستند إليه بصورة رئيسية العديد من الطرق الإحصائية، وبدونه تضيق الحلبة الواسعة لتطبيقات

الإحصاء في الحياة المعاصرة. كما أن له استنتاجات عملية عظيمة في نظرية المعاينة. حيث إنها تخبرنا بأنه عند أخذ عينات عشوائية كبيرة لمتغير ذي توزيع من أي نوع التوزيعات فإن متوسطات العينات سوف يكون لها- تقريباً- التوزيع الطبيعي. كما أنه يدخل في معظم الاختبارات الإحصائية.

ويقال للمتغير العشوائي المتصل X إنه يتبع التوزيع الطبيعي بمعالم (μ,σ^2). ويكتب على الصورة $X\sim N(,\sigma^2)$ إذا كانت دالة كثافة احتماله تعطى بالصورة الآتية:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},-\infty<x<\infty$$

حيث μ هو التوقع (الوسط)، σ هو الانحراف المعياري، $\pi=3.14159,\ e=2.71828$

واضح أن $f(x)\ \geq 0$ لجميع قيم x.

خواص التوزيع الطبيعي:

$$1-\ \int_{-\infty}^{\infty}f(x)dx=1$$

الإثبات

$$\int_{-\infty}^{\infty}f(x)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

بفرض

$$y=\frac{x-\mu}{\sigma}\Rightarrow x=\sigma y+\mu,dx=\sigma dy$$

$$\therefore\int_{-\infty}^{\infty}f(x)dx=\frac{\sigma}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}y^2}dy=\frac{2}{\sqrt{2\pi}}\int_{0}^{\infty}e^{-\frac{1}{2}y^2}dy$$

وبفرض أن

$$u=\frac{y^2}{2}\Rightarrow du=y\,dy\Rightarrow dy=\frac{1}{\sqrt{2u}}du$$

$$\therefore \int_{-\infty}^{\infty} f(x)dx = \frac{2}{\sqrt{2\pi}}\int_{0}^{\infty} \frac{e^{-u}}{\sqrt{2u}}\, du = \frac{1}{\sqrt{\pi}}\int_{0}^{\infty} u^{-\frac{1}{2}}e^{-u}\, du$$

$$= \frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\pi}}\sqrt{\pi} = 1$$

حيث $\Gamma(n)$ تسمى (دالة جاما) والتي تعرف بالصورة التالية:

$$\Gamma(n) = \int_{0}^{\infty} x^{n-1}e^{-x}\, dx$$

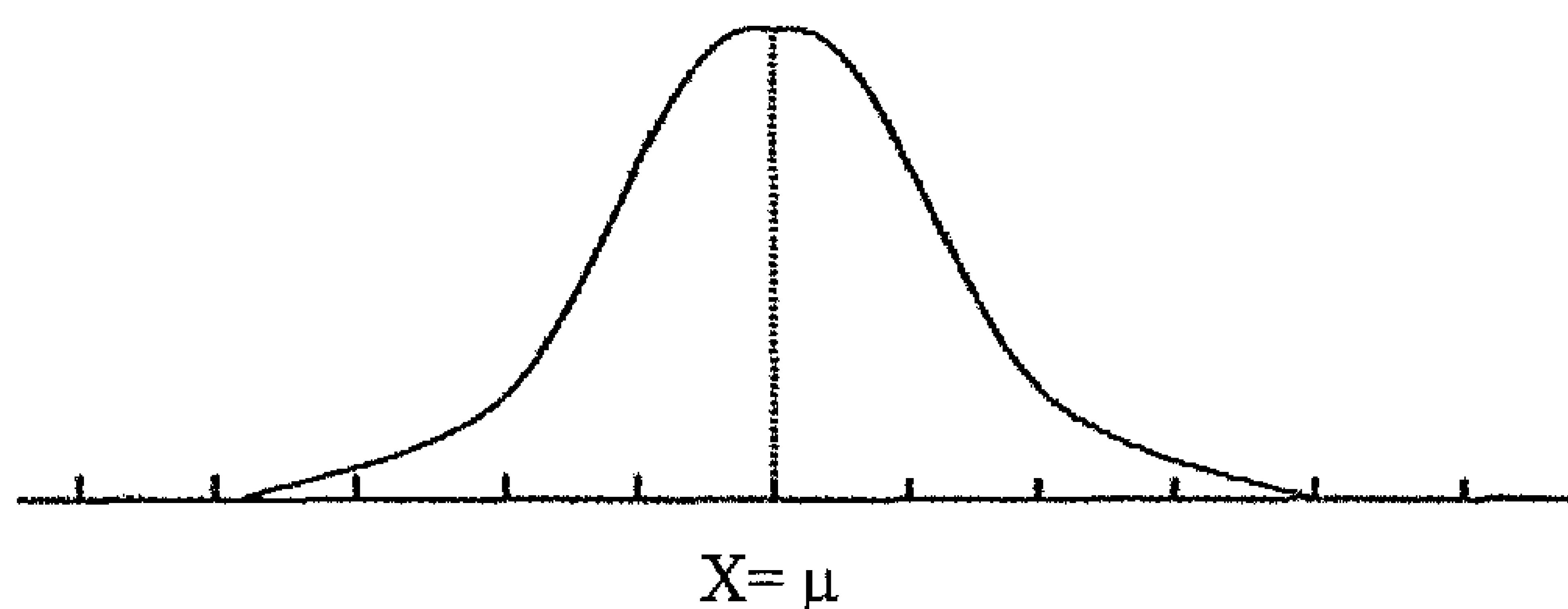
بحيث إذا كانت n عدد صحيح موجب فإن

$$\Gamma(n+1) = n\Gamma(n), \qquad \Gamma(n) = (n-1)!,$$

وأيضاً

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

وهذا يؤكد أن $f(x)$ دالة كثافة احتمال. أي أن المساحة الكلية تحت منحنى الدالة $f(x)$ يساوى الواحد الصحيح. ومنحنى التوزيع الطبيعي لها الشكل التالي:



خواص منحنى التوزيع الطبيعي:

١- المساحة الكلية تحت منحنى الدالة $f(x)$ يساوى الواحد الصحيح.

٢- منحنى دالة كثافة الاحتمال للتوزيع الطبيعي تأخذ شكل الناقوس أو الجرس.

٣- شكل المنحنى هذا يتحدد تماماً إذا علمنا قيمة الوسط μ والانحراف المعياري σ وقيمة μ تخبرنا بمركز المنحنى.

٤- منحنى التوزيع متماثل حول الوسط μ، أي أن قيمة الوسط الحسابي للتوزيع الطبيعي تساوي الوسيط، تساوي المنوال.

٥- منحني التوزيع الطبيعي يصل قيمته العظمى عند $x = \mu$ وتتوقف درجة تفرطح التوزيع على قيمة σ. فالقيمة الصغيرة لـ σ تعني أن لدينا جرساً طويلاً مدبباً بينما القيمة الكبيرة لـ σ تعني أن الجرس قصير ومفلطح.

وحيث إن متغير التوزيع الطبيعي متصل فإن هذا المتغير لا يأخذ قيمة معينة ولكنه دائما يوجد في فترة، لذلك نحسب الاحتمال بتكامل دالة كثافة الاحتمال بين حدي الفترة، أي إيجاد المساحة تحت المنحنى لدالة الكثافة.

$2 - E[X] = \mu$

الإثبات

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

بفرض أن

$$y = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma y + \mu, dx = \sigma dy$$

$$\therefore E[X] = \int_{-\infty}^{\infty} xf(x)dx = \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-\frac{1}{2}y^2} dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu) e^{-\frac{1}{2}y^2} dy$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2} dy + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy$$

وحيث إن $ye^{-\frac{1}{2}y^2}$ دالة فردية فإن:

$$\int_{0}^{\infty} y e^{-\frac{1}{2}y^2} dy = 0$$

$$\Rightarrow E[X] = 0 + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}\, dy = \frac{2\mu}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-\frac{1}{2}y^2}\, dy$$

وبفرض أن

$$u = \frac{y^2}{2} \Rightarrow du = y\, dy \Rightarrow dy = \frac{1}{\sqrt{2u}}\, du$$

$$\therefore E[X] = \frac{2\mu}{\sqrt{2\pi}} \int_{0}^{\infty} \frac{e^{-u}}{\sqrt{2u}}\, du = \frac{\mu}{\sqrt{\pi}} \int_{0}^{\infty} u^{-\frac{1}{2}} e^{-u}\, du$$

$$= \frac{\mu}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \mu$$

$3 - Var(X) = \sigma^2$

الإثبات

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\, dx$$

بفرض أن

$$y = \frac{x-\mu}{\sigma} \Rightarrow x = \sigma y + \mu, dx = \sigma dy$$

$$\therefore E[X^2] = \frac{\sigma}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma y + \mu)^2 e^{-\frac{1}{2}y^2}\, dy$$

$$= \frac{2\sigma\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y e^{-\frac{1}{2}y^2}\, dy + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma^2 y^2 + \mu^2) e^{-\frac{1}{2}y^2}\, dy$$

وحيث إن $ye^{-\frac{1}{2}y^2}$ دالة فردية فإن:

$$\int_{0}^{\infty} y e^{-\frac{1}{2}y^2}\, dy = 0$$

$$\Rightarrow E[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma^2 y^2 + \mu^2) e^{-\frac{1}{2}y^2}\, dy$$

وبفرض أن

$$u = \frac{y^2}{2} \Rightarrow du = y\,dy \Rightarrow dy = \frac{1}{\sqrt{2u}}\,du$$

$$\therefore E[X^2] = \frac{2}{\sqrt{2\pi}} \int_0^\infty (2\sigma^2 u + \mu^2) \frac{e^{-u}}{\sqrt{2u}}\,du$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^\infty u^{\frac{1}{2}} e^{-u}\,du + \frac{\mu^2}{\sqrt{\pi}} \int_0^\infty u^{-\frac{1}{2}} e^{-u}\,du$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) + \frac{\mu^2}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right) = \frac{2\sigma^2}{\sqrt{\pi}} \frac{1}{2}\,\Gamma\left(\frac{1}{2}\right) + \frac{\mu^2}{\sqrt{\pi}} \Gamma\left(\frac{1}{2}\right)$$

$$= \sigma^2 + \mu^2$$

$$\therefore Var(X) = E[X^2] - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

ولحساب الاحتمالات على التوزيع الطبيعي فإنه يجب إجراء تكامل على دالة الكثافة بحيث

$$P[a \le X \le b] = \int_a^b f(x)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\,dx$$

وهذا التكامل صعب تعيينه لذا كان لزاما إيجاد حل لكيفية حساب الاحتمالات للتوزيع الطبيعي.

٩-٥ التوزيع الطبيعي القياسي (Standard Normal Distribution)

يمكن تحويل كل توزيع طبيعي إلى توزيع آخر يسمى التوزيع الطبيعي القياسي (المعياري) وذلك لحساب الاحتمالات بسهولة من خلال جداول تسمى جداول التوزيع الطبيعي المعياري، فإذا كانت $\mu = 0,\ \sigma^2 = 1$ في التوزيع الطبيعي فإن التوزيع في هذه الحالة يسمى التوزيع الطبيعي المعياري. ومن ثمّ فإن المتغير العشوائي Z يقال أنه يتبع التوزيع الطبيعي القياسي (المعياري) $Z \sim N(0,1)$ إذا كانت دالة كثافة احتماله تعطى بالصورة التالية:

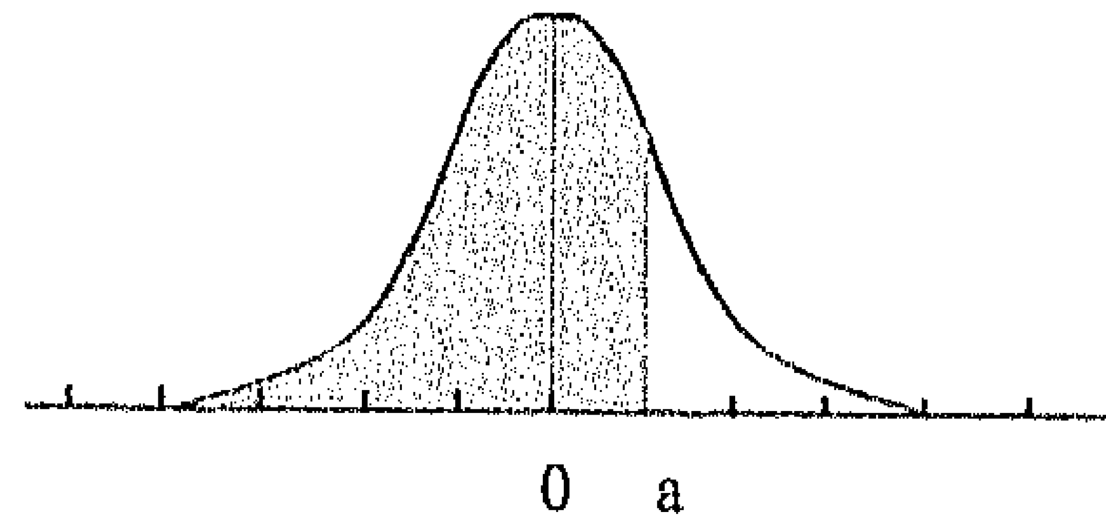
$$\varphi(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2}, -\infty < z < \infty$$

ودالة التوزيع للمتغير Z تعطى بالصورة التالية:

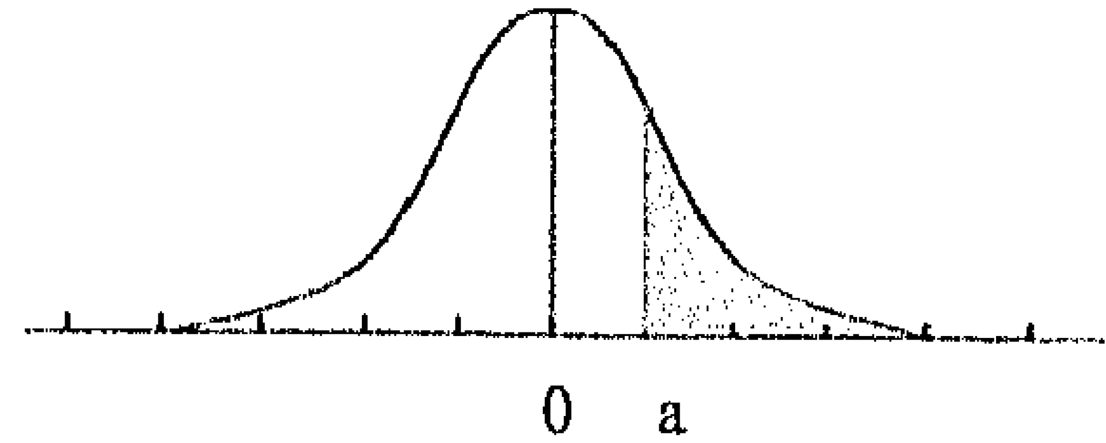
$$\Phi(z) = P[Z \leq z] = \int_{-\infty}^{z} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\, dt$$

فإذا كان a, b ثوابت غير سالبة فإن:

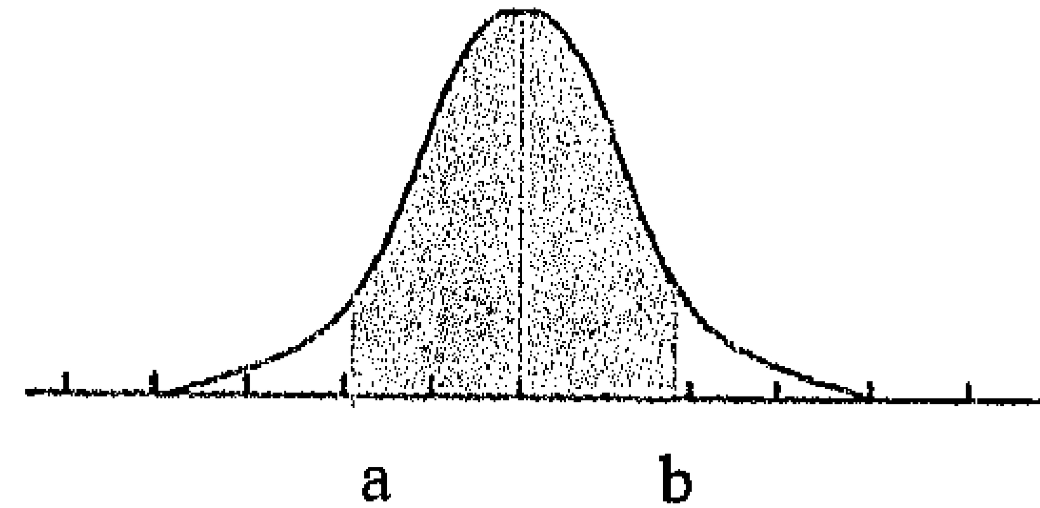
$$(i) P[Z \leq a] = \Phi(a)$$



$$(ii) P[Z > a] = 1 - P[Z \leq a] = 1 - \Phi(a)$$



$$(iii) P[a \leq Z \leq b] = \Phi(b) - \Phi(a)$$



$$(iv)\ \Phi(-a) = 1 - \Phi(a)$$

ومن حيث أن المعنى الهندسي للتكامل المحدود هو المساحة تحت منحنى الدالة في فترة التكامل فإنه يوجد جدول يحتوى قيم دالة التوزيع $\Phi(z)$ لقيم z الموجبة ويسمى جدول التوزيع الطبيعي المعياري. ومن ثمَّ يمكن حل مشكلة تعيين الاحتمالات على التوزيع الطبيعي بإمكانية تحويل المتغير X الذي يتبع التوزيع الطبيعي إلى متغير Z يتبع التوزيع الطبيعي القياسي بإجراء تحويله عليه كما في النظرية التالية.

نظرية (٩-١)

إذا كان X متغيراً عشوائياً يتبع التوزيع الطبيعي بمتوسط μ وتباين σ^2 فإن المتغير $Z = \frac{X-\mu}{\sigma}$ يتبع التوزيع الطبيعي المعياري بمتوسط 0 وتباين 1، أي أن

$$X \sim N(\mu, \sigma^2) \Longrightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

البرهان

$$P[Z \le z] = P\left[\frac{X-\mu}{\sigma} \le z\right] = P[X \le \sigma z + \mu]$$

لكن $X \sim (\mu, \sigma^2)$ لذا فإن

$$P[Z \le z] = P[X \le \sigma z + \mu] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\sigma z+\mu} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

وبفرض أن

$$t = \frac{x-\mu}{\sigma} \Longrightarrow dx = \sigma dt, \qquad t: -\infty \longrightarrow z$$

$$\therefore P[Z \le z] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} \sigma dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{1}{2}t^2} dt = \Phi(z)$$

وهذه دالة التوزيع التراكمية للتوزيع الطبيعي المعياري $N(0,1)$

ويتم تعيين قيمة $\Phi(a)$ من جدول التوزيع الطبيعى القياسى عن طريق تقسيم العدد a لمجموع عددين أحدهما العدد الحقيقي الذي يحتوى على رقم واحد فقط بعد العلامة والعدد الثاني هو عدد حقيقي يحتوى على رقمين بعد العلامة، فعلى سبيل المثال إذا كانت $a = 1.96$ فإنه يكتب على الصورة التالية:

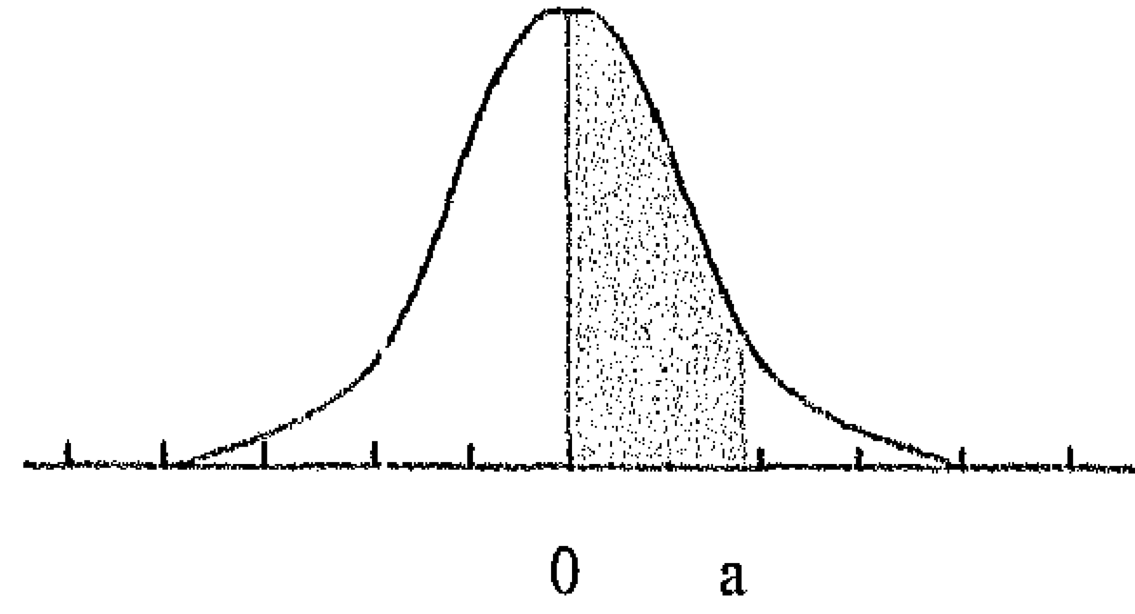
$$a = 1.96 = 1.9 + 0.06$$

ومن جدول التوزيع الطبيعي القياسي نبحث في بداية الصفوف عن الصف الذي يبدأ بالقيمة 1.9 والعمود الذي يبدأ بالقيمة 0.06 ونقطة تلاقيهم (تقاطعهم) هي قيمة $\Phi(1.96)$ وهى 0.975، ويجب ملاحظة أن جدول دالة التوزيع أقل قيمة في الجدول هي 0.5 أي أن $\Phi(0.0) = 0.5$

ويجب ملاحظه أن

$$\Phi(a) = P[Z \le a] = \int_{-\infty}^{a} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{1}{2}t^2}\, dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{1}{2}t^2}\, dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{a} e^{-\frac{1}{2}t^2}\, dt = 0.5 + P[0 \le Z \le a]$$



وأحيانا يكون لدينا جدول يعطى قيمة الاحتمالات ابتدأ من الصفر حتى القيمة a ومن هذا الجدول يمكن تعيين قيمة $\Phi(a)$ عن طريق اضافة 0.5 للقيمة التي سنحصل عليها من الجدول بنفس طريقة البحث السابق ذكرها.

مثال (٩-٥)

أوجد قيمة $\Phi(z)$ إذا كانت z = 1.82 ,z = 2.58 , z = - 0.47

الحل

باستخدام جدول التوزيع الطبيعي القياسي فإن

$\Phi(1.82) = P[Z \le 1.82] = 0.96562$

$\Phi(2.58) = P[Z \le 2.52] = 0.99506$

$\Phi(-0.47) = 1 - \Phi(0.47) = 1 - 0.6808 = 0.3192$

مثال (٩-٦)

إذا كان X متغيراً عشوائياً بحيث X ~ N(50, 25) فأوجد الاحتمالات التالية

(i)$P[\,40 < X < 60]$

(ii)$P[\,X > 60]$

(iii)$P[\,55 < X < 65]$

(iv)$P[\,X \leq 45\,]$

الحل

من المعطيات فإن $\mu = 50\,,\ \sigma = 5$

(i) $P[\,40 < X < 60] = P\left[\frac{40-50}{5} < \frac{X-\mu}{\sigma} < \frac{60-50}{5}\right]$

$$= P[-2 < Z < 2] = \Phi(2) - \Phi(-2)$$

$$= \Phi(2) - [1 - \Phi(2)] = 2\Phi(2) - 1 = 0.9544$$

(ii) $P[\,X > 60] = P\left[\frac{X-\mu}{\sigma} > \frac{60-50}{5}\right] = P[Z > 2]$

$$= 1 - \Phi(2) = 1 - 0.9772 = 0.0228$$

(iii) $P[\,55 < X < 65] = P\left[\frac{55-50}{5} < \frac{X-\mu}{\sigma}, \frac{65-50}{5}\right] = P[1 < Z < 3]$

$$= \Phi(3) - \Phi(1) = 0.9987 - 0.8413 = 0.1574$$

(iv) $P[\,X \leq 45] = P\left[\frac{X-\mu}{\sigma} \leq \frac{45-50}{5}\right] = P[Z \leq -1]$

$$= \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$$

مثال (٩-٧)

إذا كان X متغيراً عشوائياً يتبع التوزيع الطبيعي $N(\,18, 6.25)$ فأوجد قيمة الثابت K إذا كان $P[\,X > K] = 0.1539$

الحل

بما أن المتغير العشوائي X يتبع التوزيع الطبيعي حيث $\mu = 18\,,\quad \sigma = 2.5$ فإن

$$P[X > k] = 0.1539 \Longrightarrow P\left[\frac{X-\mu}{\sigma} > \frac{k-\mu}{\sigma}\right] = 0.1539$$

$$\Longrightarrow P\left[Z > \frac{k-\mu}{\sigma}\right] = 0.1539 \Longrightarrow 1 - \Phi\left(\frac{k-\mu}{\sigma}\right) = 0.1539$$

$$\Rightarrow \Phi\left(\frac{k-\mu}{\sigma}\right) = 1 - 0.1539 = 0.8461$$

ومن جدول التوزيع الطبيعي القياسي نبحث عن القيمة 0.8461 فنجدها تقع في الصف الذي يبدأ بالرقم 1.0 وأيضا تقع في العمود الذي يبدأ بالقيمة 0.02 لذا فإن الرقم المطلوب هو 1.02 وبالتالي فإن

$$\Rightarrow \frac{k-\mu}{\sigma} = \frac{k-18}{2.5} = 1.02 \;\Rightarrow k = 1.02(2.5) + 18 = 20.55$$

مثال (٩-٨)

اذا علم أن المتغير X يمثل أطوال ورقة نبات ما (مَقِيسة بالمليمترات) الذى يتوزع طبيعياً بحيث إن $X \sim N(132,100)$ والمطلوب حساب نسبة الأوراق التى طولها ما بين 128 مم، 142 مم وكذلك عددها فى عينة مكونة من 500 ورقة.

الحل:

بما أن $X \sim N(132,100)$ فإن

$$Z = \frac{X-\mu}{\sigma} = \frac{X-132}{10} \sim N(0,1)$$

النسبة المطلوبة تعنى الاحتمال وهى

$$P[128 \le X \le 142] = P\left[\frac{128-132}{10} \le \frac{X-\mu}{\sigma} \le \frac{142-132}{10}\right]$$

$$= P[-0.4 \le Z \le 1]$$

$$= \Phi(1) - \Phi(-0.4) = \Phi(1) - 1 + \Phi(0.4) = 0.4967$$

أى أن نسبة الأوراق داخل الفترة (128, 142) مم تقريباً هى 50% وعدد الأوراق المطلوبة من مجموعة 500 ورقة هى

$$0.4967 \times 500 = 248.35 \cong 249$$

٩-٦ توزيع t (Student`s t Distribution)

من التوزيعات التى لها دور لكن أقل قليلاً من التوزيع الطبيعى والتى تستخدم فى الاستدلال الإحصائي فى نظرية التقدير واختبارات الفروض الإحصائية، ما يعرف بتوزيع t والمعروف أيضاً بتوزيع الطالب (Student`s

distribution). ويقال إن المتغير العشوائى T يتبع توزيع t بمعلم ν إذا كانت دالة كثافته الاحتمالية على الصورة التالية:

$$f(t) = \frac{1}{\sqrt{\nu}B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu}{2}-\frac{1}{2}} \quad ,-\infty < t < \infty$$

وتسمى ν ب(درجة الحرية) لتوزيع t والتى تعطى بالصورة $\nu = n-1$ وأيضاً $\beta(n,m)$ تسمي (دالة بيتا) والتي تعرف بالصورة التالية:

$$\beta(n,m) = \int_a^b (x-a)^{n-1}(b-x)^{m-1}dx$$

حيث إن

$$\beta(n,m) = \frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m)}.$$

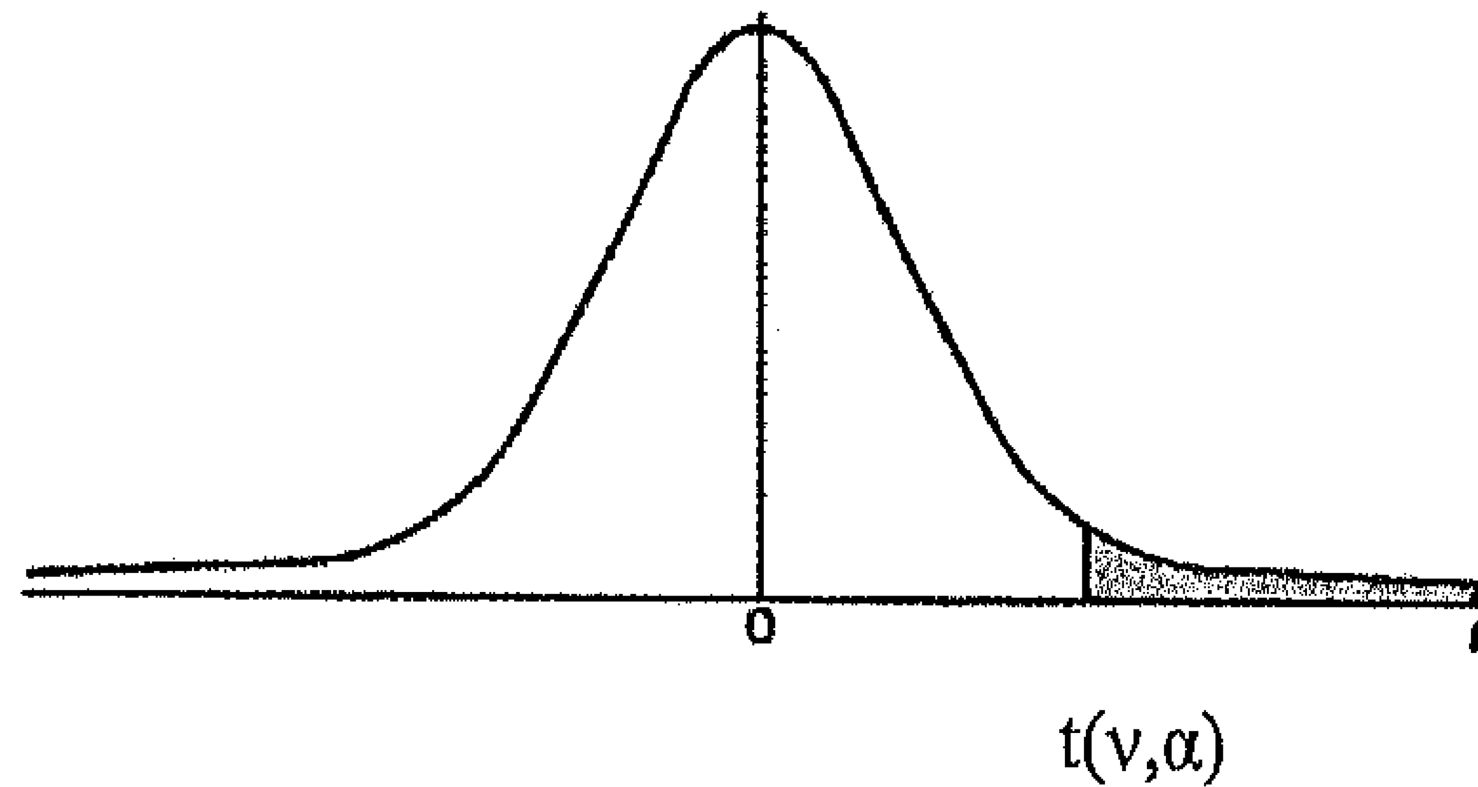
خواص توزيع t :

(أ) يوجد عدد لانهائي من توزيعات t نتيجة لتغيير قيمة ν

(ب) منحنى توزيع t يشبة الناقوس ومتماثل حول المحور $t = 0$

(ج) التوقع لتوزيع t هو $E(T)=0$ والتباين $Var(T) = \frac{\nu}{\nu-2}$ حيث $\nu > 2$

ومن المعلوم أنه لتعيين الاحتمالات للمتغير المتصل نقوم بتكامل دالة الكثافة الاحتمالية على الفترة التى نرغب فى تعيين الاحتمال فيها. ومن الواضح أن عملية التكامل على توزيع t يصعب تعيينة لذا فلحساب قيمة الاحتمال حول المتغير T يوجد جدول يعطي تلك الاحتمالات والتى تعرف بالمساحات تحت منحنى الدالة ويعطى الجدول القيمة $P[T \geq t_{\nu,\alpha}] = \alpha$ والتى توضح بالمنطقة المظللة على منحني توزيع t في الشكل التالي:



t(ν,α)

حيث $t_{v,\alpha}$ هى نقطة على الخط الأفقى، α المساحة الموجودة على يمين النقطة،ν درجة الحرية التى يتميز بها التوزيع. فإذا أعطيت اثنين من الثلاثة $\nu, \alpha, t_{v,\alpha}$, فيمكن إيجاد القيمة الثالثة،حيث توجد α فوق الجدول وتوجد ν على يسار الجدول وتوجد $t_{v,\alpha}$ داخل الجدول.

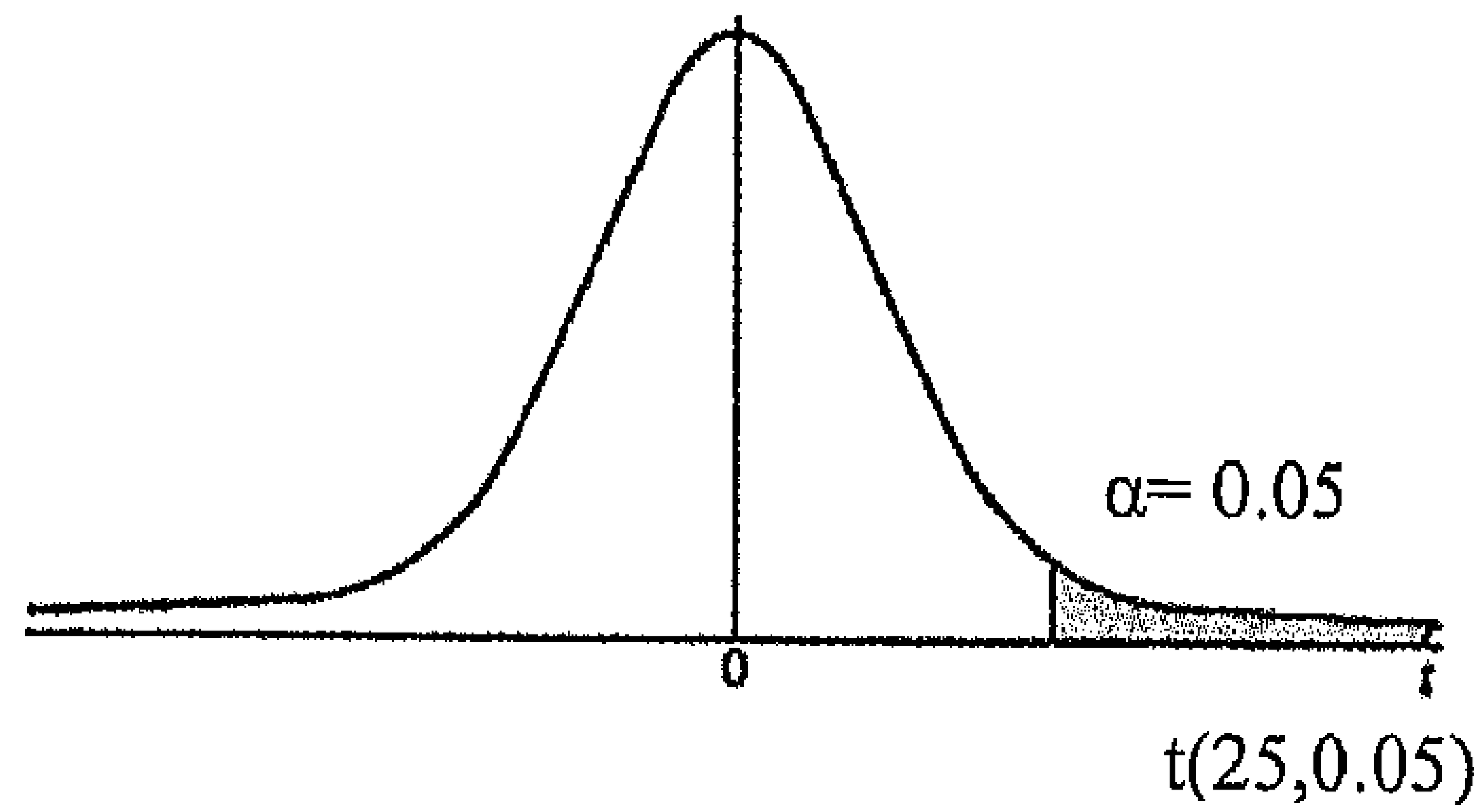
مثال (٩-٩)

(أ) احسب قيمة $t_{v,\alpha}$ التى تجعل $P[T \geq t_{\nu,\alpha}] = 0.05$ علما بأن ν=25

(ب) أوجد الاحتمال $P[t_{\nu,0.95} \leq T \leq t_{\nu,0.01}]$ علما بان ν = 25

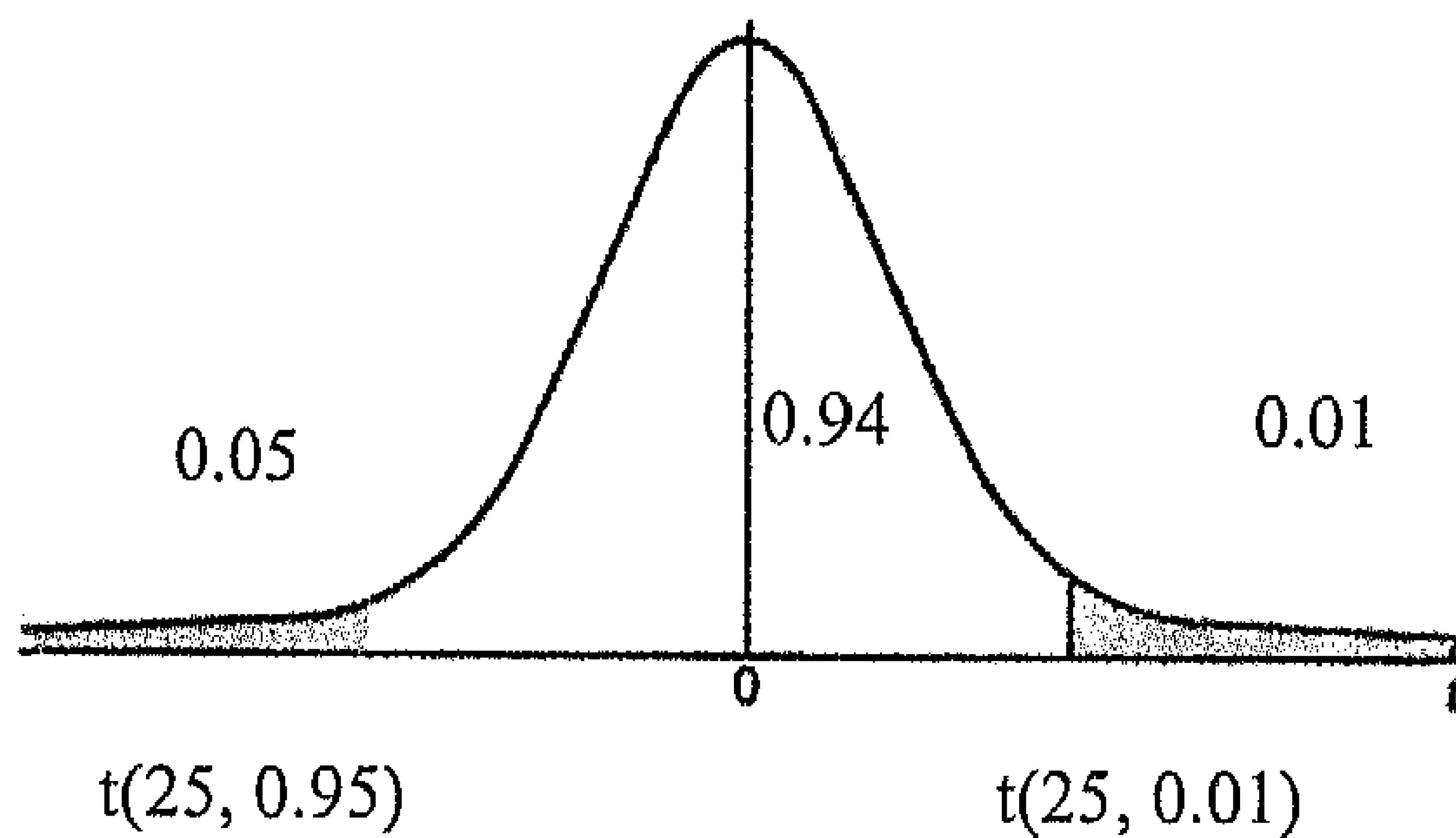
الحل:

$P[T \geq t_{\nu,\alpha}] = 0.05$



من جدول توزيع t حيث معلوم لدينا كلاً من $\nu = 25, \alpha = 0.05$ لذا سوف نبحث عن $t_{25,0.05}$ وتوجد α اعلى الجدول فنبحث عن القيمة 0.05 وتوجد درجة الحرية على يسار الجدول فنبحث عن 25 وتتقابل القيمتان عند القيمة 1.708 وهى قيمة t المطلوبة.

$P[t_{\nu,0.95} \leq T \leq t_{\nu,0.01}]$



قيمة هذا الاحتمال هو المساحة الموجودة أسفل منحنى توزيع t والمحصورة بين النقطتين $t(25, 0.01), t(25, 0.95)$ ومن الشكل المقابل نجد أن تلك المساحة هى 0.94

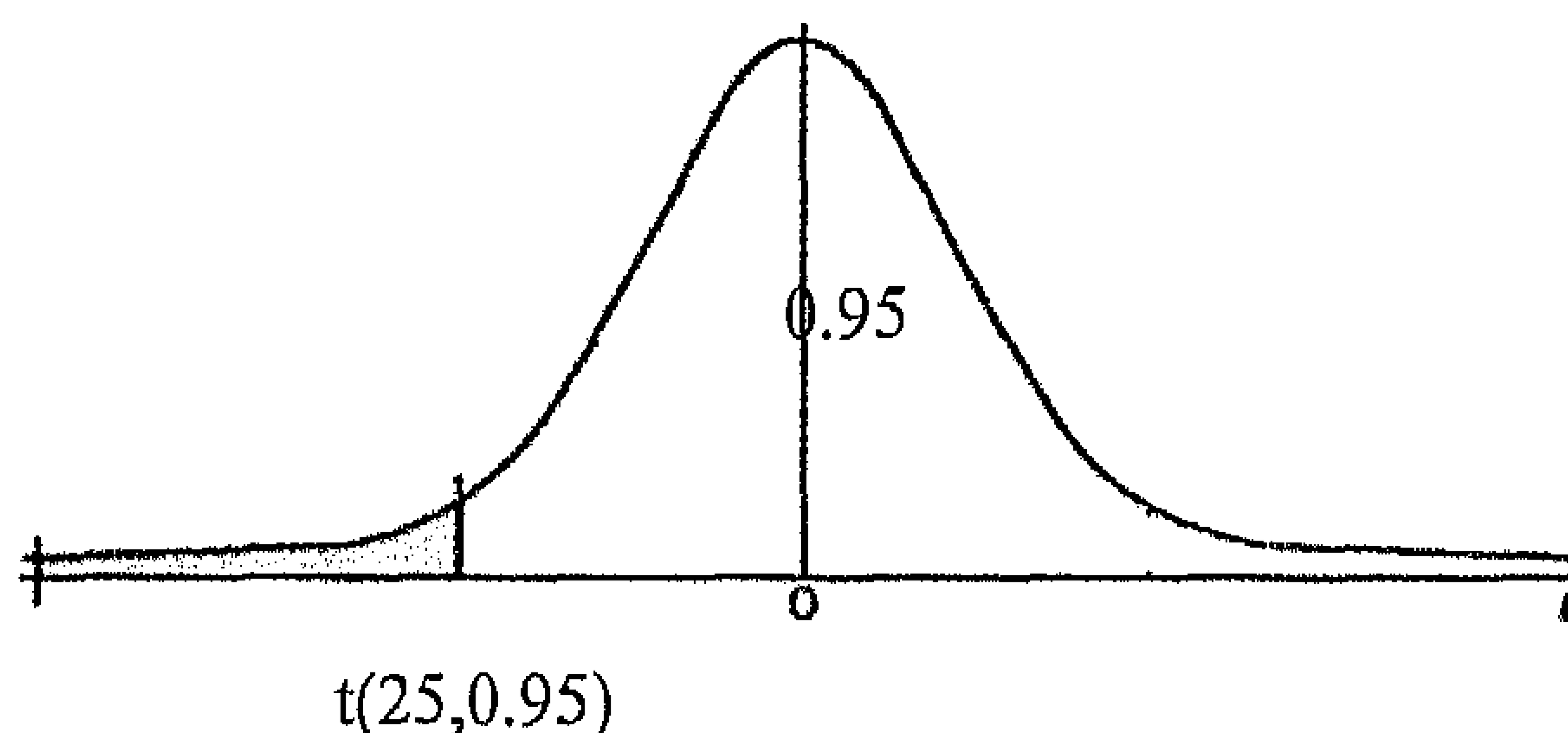
ملحوظة: بما أن منحنى توزيع t متماثل حول المحور $t = 0$ لذا فإن

$$t_{v,\alpha} = -t_{v,1-\alpha}$$

مثال(٩-١٠)

أوجد قيمة $t(25, 0.95)$

الحل:



نلاحظ أن القيمة المطلوبة يوجد على يمينها مساحة مقدارها 0.95 وبالتالى فهى تقع فى الجزء السالب من المحور الأفقى فهى قيمة سالبة (لو بحثنا فى الجدول عن قيمة α=0.95 فلن نجدها وهذا يدل على ان القيمة سالبة) والجدول لا يعطى إلا القيم الموجبة، لذا سوف نستخدم العلاقة السابقة:

$$t_{v,\alpha} = -t_{v,1-\alpha}$$

حيث

$$t_{25,0.95} = -t_{25,0.05} = -1.708$$

٩-٧ توزيع مربع كاي (Chi-square Distribution)

توزيع مربع كاي يرمز له بالرمز χ^2 وتقرأ (كاي سكوير). وهو من التوزيعات التى تؤدي دوراً مهما فى الحياة والتطبيقات العملية. ويستخدم فى الاختبارات الإحصائية والتقدير، فيقال للمتغير العشوائي χ^2 إنه يتبع توزيع مربع كاي بمعلمة v إذا كانت دالة الكثافة الاحتمالية له تعطى بالصورة التالية:

$$f(\chi^2) = \frac{1}{2\Gamma\left(\frac{v}{2}\right)} \left(\frac{\chi_v^2}{2}\right)^{\frac{v}{2}-1} e^{-\frac{\chi_v^2}{2}}, \;\; 0 < \chi_v^2 < \infty$$

وتسمى ν بدرجة الحرية لتوزيع مربع كاي حيث $\nu = n - 1$.

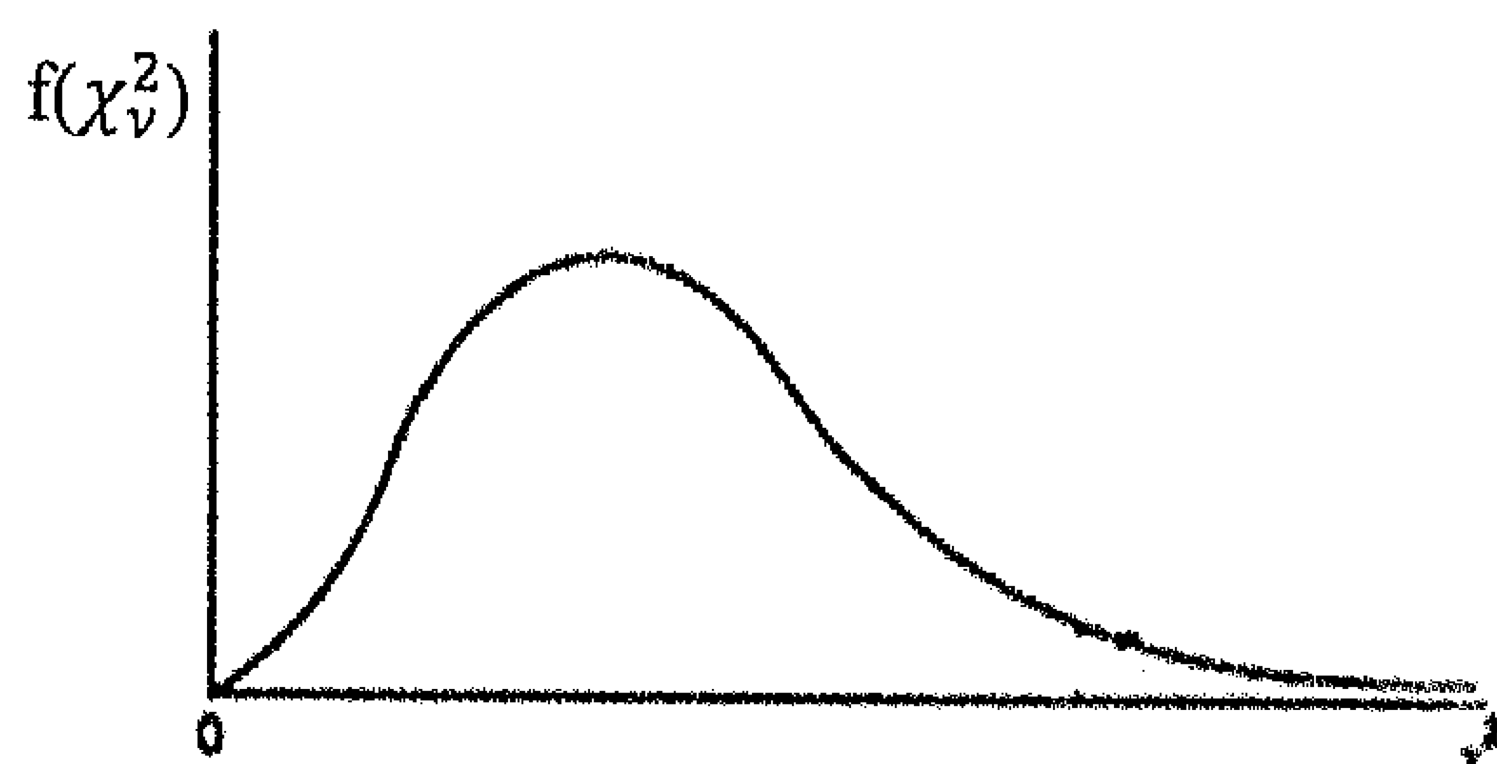
خواص توزيع مربع كاي:

(أ) يوجد عدد لانهائي من توزيعات مربع كاي تعتمد على قيمة ν

(ب) المتغير χ^2 يأخذ قيما موجبة فقط.

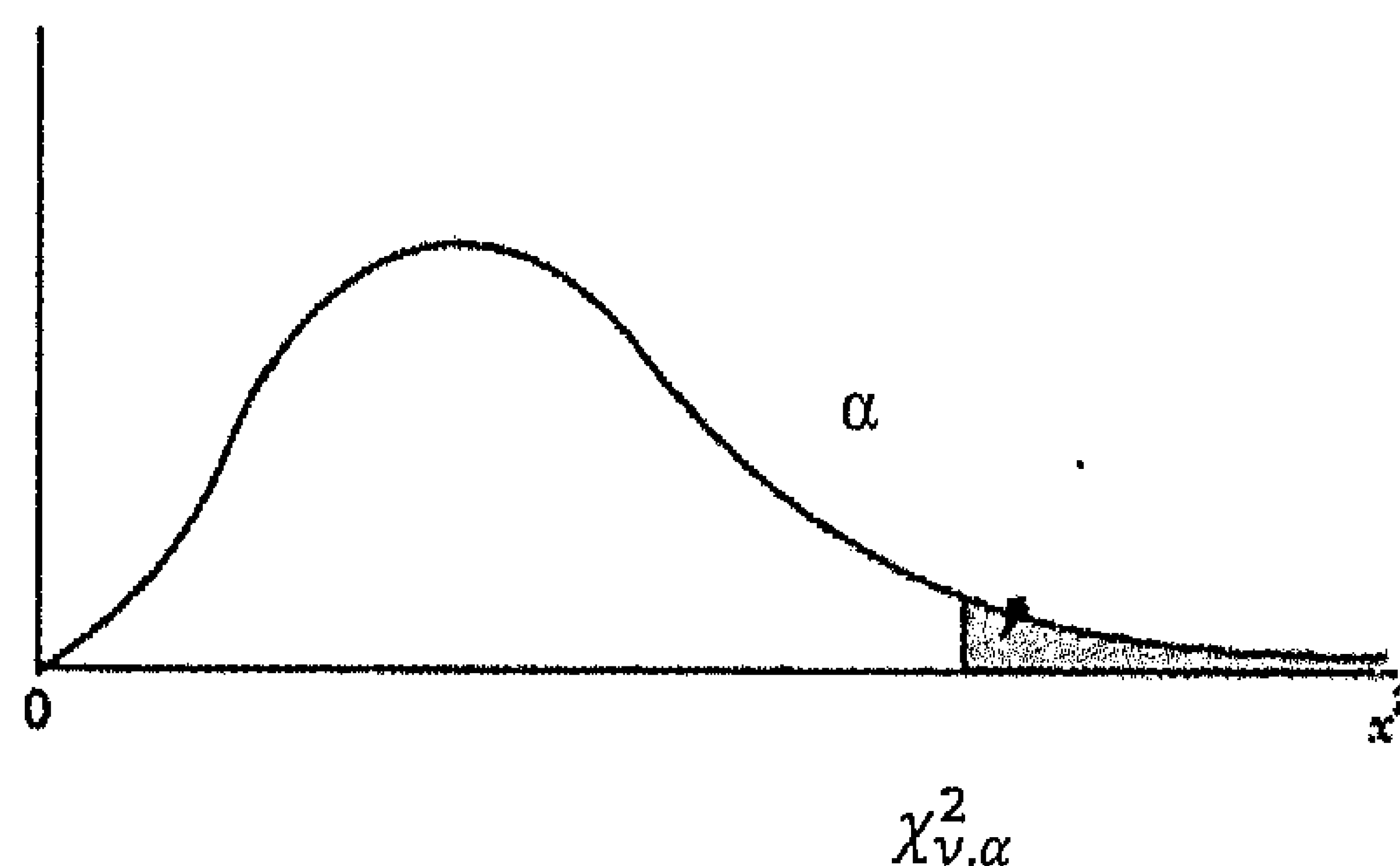
(ج) التوقع لتوزيع مربع كاي هو $E(\chi_\nu^2) = \nu$ والتباين $Var(\chi_\nu^2) = 2\nu$

ومنحنى توزيع مربع كاي له الشكل التالي:



ويوجد جدول لتوزيع مربع كاي يعطى قيمة الاحتمال التالي:

$$P\left[\chi_v^2 \geq \chi_{v,\alpha}^2\right] = \alpha$$



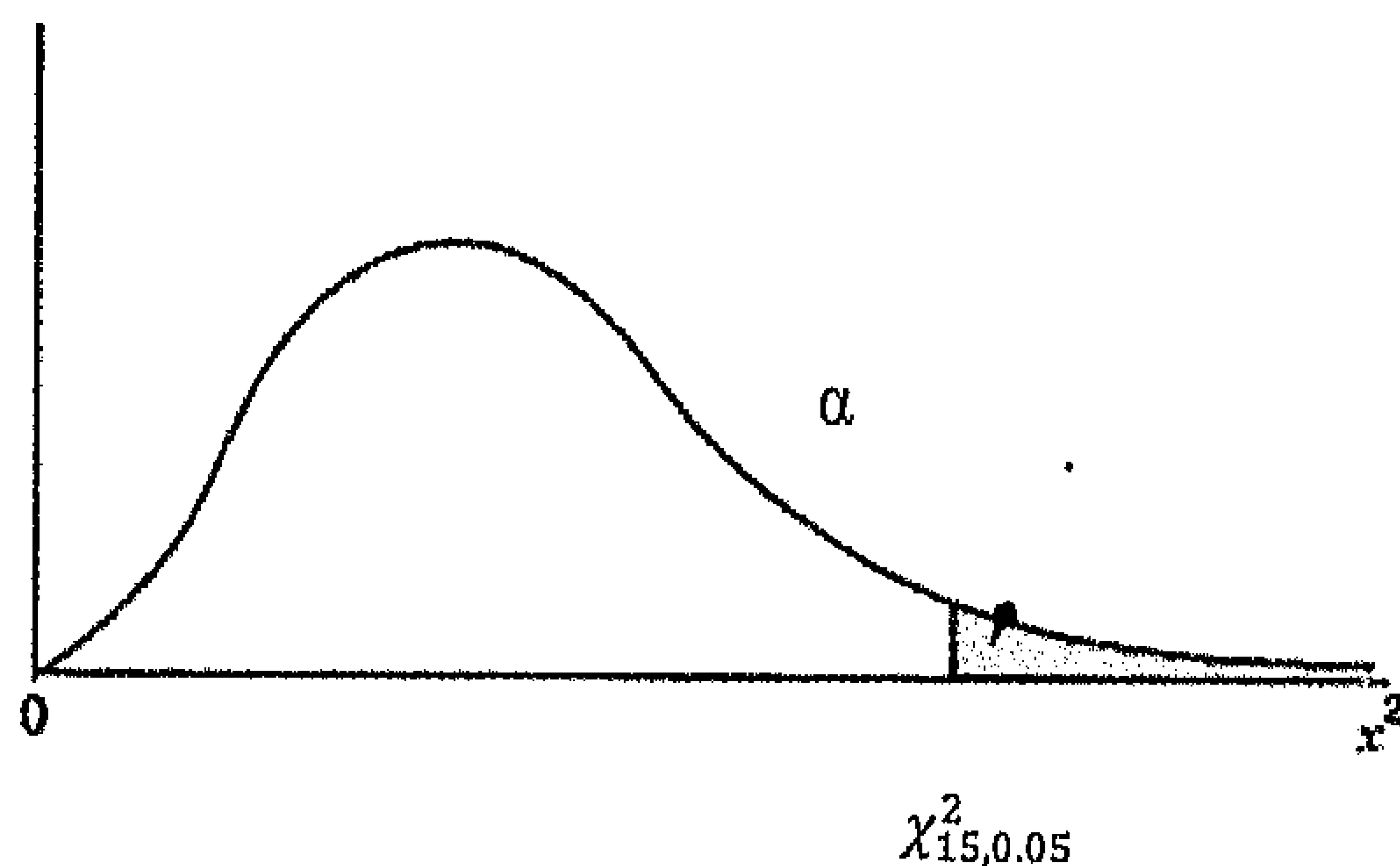
حيث $\chi^2_{\alpha,\nu}$ هى نقطة على الخط الأفقى، α المساحة الموجودة على يمين النقطة،ν درجة الحرية التى يتميز بها التوزيع. فإذا أُعطيت اثنين من الثلاثة $\nu,\ \alpha,\ \chi^2_{\alpha,\nu}$ فيمكن إيجاد القيمة الثالثة. حيث توجد α فوق الجدول وتوجد ν على يسار الجدول وتوجد $\chi^2_{\alpha,\nu}$ داخل الجدول.

مثال(٩-١١)

أوجد قيمة $\chi^2_{15,0.05}$ والتى تجعل $P[\chi^2_v > \chi^2_{15,0.05}] = 0.05$

الحل:

$$P[\chi^2_v > \chi^2_{15,0.05}] = 0.05$$



ومن جدول التوزيع نجد أن $\chi^2_{15,0.05} = 24.996$

أسئلة وتمارين (٩)

١- اذكر خواص التوزيع الطبيعي، وارسم منحنى دالة الكثافة له.

٢- اذكر خواص التوزيع الآسي.

٣- أثبت أن التوقع الرياضي للتوزيع الطبيعي القياسي هو 0 والتباين 1.

٤- ما الفرق بين التوزيع الطبيعي والتوزيع الطبيعي القياسي؟

٥- كيف يمكن التحويل من التوزيع الطبيعي إلى التوزيع الطبيعي القياسي؟

٦- المتغير العشوائي X يمثل الوزن مقدراً بالرطل لسمكة من نوع السالمون الكبير والتي تصطاد عند مصب نهر معين. بفرض أن X له التوزيع الطبيعي بمتوسط 30 رطلاً وانحراف معياري 6 أرطال. تم الحصول على سمكة من هذا النوع أوجد احتمال أن يكون وزنها:

(أ) على الأقل 41 رطلاً.

(ب) ألا يزيد وزنها عن 45 رطلاً.

(ج) أن يتراوح بين 20 رطلاً ، 40 رطلاً.

٧- شركة تصنع مصابيح كهربائية من نوع معين. وكان عمر المصباح (بالساعة) له التوزيع الطبيعي بمتوسط 800 وانحراف معياري 40 أوجد احتمال أن:

(أ) يعمل المصباح ما بين 773، 834 ساعة.

(ب) يعمل المصباح أكثر من 900 ساعة.

(ج) لا يعمل المصباح أكثر من 800 ساعة.

٨- أوجد $P[1 \leq X \leq 2]$ للمتغير العشوائى X عندما

(أ) X له التوزيع الطبيعي بمتوسط 1 وتباين 1

(ب) X له التوزيع الآسي $\lambda=1$

(ج) X له التوزيع المنتظم على الفترة $(\frac{1}{2},\frac{3}{2})$

٩- إذا فرض أن آلة وضعت في الخدمة وكان X متغيراً عشوائياً يتبع التوزيع الأسي ببارامتر λ ، ويمثل الوقت اللازم حتى تتعطل هذه الآلة عن العمل. أثبت أنه مهما كانت a فإن الاحتمال $P[X > a + b \,|X > a]$ لا يعتمد على a.

١٠- أثبت أن التوقع الرياضي للتوزيع الطبيعي المعياري يساوى صفراً والتباين يساوى الواحد الصحيح.

١١- إذا كان X متغيراً عشوائياً له دالة كثافته على الصورة:

$$f(x) = k\, e^{-\frac{1}{2}\left(\frac{x-\alpha}{\beta}\right)^2}, -\infty < x < \infty$$

أوجد قيمة الثابت K

١٢- محطة أتوبيسات يصل إليها أتوبيس كل 15 دقيقة بدءاً من الساعة السابعة صباحاً، إذا وصل أحد الركاب إلى المحطة بين الساعة السابعة والسابعة والنصف فأوجد احتمال أنه سينتظر: أقل من 5 دقائق ليركب الأتوبيس، أكثر من 10 دقائق ليركب الأتوبيس.

١٣- إذا كان $X \sim N(\mu, \sigma^2)$ احسب $P[\mu - 2\sigma \le x \le \mu + 2\sigma]$

١٤- إذا كان X متغيراً عشوائياً دالة كثافته الاحتمالية هي

$$f(x) = cx^2(1-x), \qquad 0 \le x \le 1$$

(أ) أوجد قيمة الثابت c

(ب) عين دالة التوزيع للمتغير العشوائي X

(ج) احسب قيمة الاحتمال $P[X \le 0.6]$

(د) أوجد متوسط وتباين هذا التوزيع.

١٥- إذا كان عمر نوع معين من أجهزة الهاتف المحمول التي تنتجها إحدى الشركات تتبع التوزيع الطبيعي بمتوسط 10 سنوات وانحراف معياري سنتين. إذا قررت إدارة الشركة المصنعة وضع عرض دعائي لمنتجها بحيث تستبدل أي جهاز يتعطل أثناء فترة الضمان مجاناً. فكم يجب أن تكون فترة الضمان التي تعطى للزبون إذا رغبت الشركة أن لا تزيد كمية الهواتف المستبدلة عن 1% من الهواتف التي تنتجها الشركة؟

١٦- إذا كان متوسط أعمار المصابيح التي ينتجها أحد المصانع هو 1500 ساعة. فما هو احتمال

(أ) أن يعيش أحد المصابيح أكثر من 3000 ساعة.

(ب) أن يحترق أحد المصابيح خلال 150 ساعة.

(ج) أن يعيش المصباح أكثر من 5000 ساعة علماً بأنه قد عمل أكثر من 2000 ساعة.

١٧- إذا كانت العلاقة بين الدخل X والاستهلاك Y لأسرة في مدينة ما هي $Y = 0.75\,X + 20$ وكان الدخل يتبع التوزيع الطبيعي بمتوسط 3000 ريال وانحراف معياري 600 ريال، اخترنا عشوائياً أسرة فما احتمال أن:

(أ) يزيد دخلها عن 3900 ريالاً،

(ب) يقل استهلاكها عن 2490.5 ريالاً؟

١٨- إذا كان X متغير عشوائي له التوزيع المنتظم على الفترة (0, 10) فأوجد

(أ) الاحتمالات التالية:

$$P[X < 3], \qquad P[X > 6], \qquad P[3 < X < 8]$$

(ب) التوقع الرياضي والتباين للمتغير العشوائي X

١٩- بفرض أن طول المكالمة التليفونية بالدقيقة تتبع التوزيع الأسى بمعلمة 0.1 إذا وصل صديق إليك وأنت تجرى مكالمة تليفونية فما هو احتمال أنه سينتظرك

(أ) أكثر من 10 دقائق.

(ب) ما بين 10 إلى 20 دقيقة.

(ج) أقل من 15 دقيقة.

٢٠- بفرض أن عدد الأميال التي تسيرها السيارة قبل تلف البطارية يتبع التوزيع الأسي بمتوسط 10000 ميلاً. إذا أراد شخص أن يسافر لمسافة 5000 ميل فما هو احتمال أنه سيكمل سفره بدون استبدال البطارية؟

٢١- في محطة قطارات إذا كانت القطارات تتحرك من الرصيف A والتي تصل للمحطة كل 15 دقيقة ابتداء من الساعة 7 صباحاً، أو من الرصيف B والتي تصل للمحطة كل 15 دقيقة ابتدأ من الساعة 7:05 صباحاً. إذا وصل مسافر ما للمحطة في الفترة الزمنية من الساعة 7 إلى 8 والتي تتبع التوزيع المنتظم وسوف يركب أول قطار سيصل للمحطة فما احتمال أنه سوف يتجه للرصيف A؟

٢٢- إذا وصلت لمحطة أتوبيسات في الساعة العاشرة وكان معلوماً أن زمن وصول الأتوبيسات يتبع التوزيع المنتظم في الفترة من 10 حتى 10:30

(أ) فما هو احتمال أنك ستنتظر الأتوبيس أكثر من 10 دقائق.

(ب) إذا وصلت الساعة 10:15 ولم يصل الأتوبيس بعد، فما احتمال أنك ستنتظر على الأقل 10 دقائق إضافية.

٢٣- إذا كان زمن إصلاح الآلة المتعطلة بالساعات تتبع التوزيع الأسي بمتوسط ساعتين. فأوجد

(أ) احتمال أن يزيد وقت الإصلاح عن ساعتين.

(ب) الاحتمال الشرطي لأن يكون وقت إصلاح الآلة سيأخذ على الأقل 10 ساعات. علماً بأنه قد زاد عن 9 ساعات؟

٢٤- إذا كان X متغيراً عشوائياً يتبع التوزيع الأسي بمتوسط $\frac{1}{\lambda}$ فأثبت أن

$$E[X^k] = \frac{k!}{\lambda^k}, \quad k = 1, 2, 3, \dots$$

ومنها أوجد قيمة $E[X^4]$

٢٥- إذا كان X متغيراً عشوائياً دالة كثافة احتماله هي

$$f(x) = ce^{-3x}, \qquad 0 \le x \le \infty$$

(أ) أوجد قيمة الثابت c

(ب) عين دالة التوزيع للمتغير العشوائي X

(ج) احسب قيمة الاحتمال $P[X \le 0.6]$

(د) احسب قيمة الاحتمال $P[0.3 \le X \le 1.3]$

(و) أوجد متوسط وتباين هذا التوزيع.

٢٦- باستخدام جدول توزيع t أوجد القيم التالية:

$$t(6, 0.05), \; t(12, 0.025), \; t(10, 0.99), t(24, 0.005), -t(8, 0.995)$$

٢٧- أوجد قيمة $t_{v,\alpha}$ عندما $\nu = 15$ في الحالات التالية:

(i) $P[t_\nu \ge t_{v,\alpha}] = 0.005$

(ii) $P[t_\nu \ge t_{v,\alpha}] = 0.01$

(iii) $P[t_\nu \ge t_{v,\alpha}] = 0.025$

(iv) $P[t_{\nu,0.99} \le t_\nu \le t_{v,0.05}]$

٢٨- إذا كانت $v = 10$ فأوجد قيمة $\chi^2_{v,\alpha}$ فى الحالات التالية:

(i) $P\left[\chi^2_v \geq \chi^2_{v,\alpha}\right] = 0.005$

(ii) $P\left[\chi^2_v \geq \chi^2_{v,\alpha}\right] = 0.01$

(iii) $P\left[\chi^2_v \geq \chi^2_{v,\alpha}\right] = 0.025$

(iv) $P\left[\chi^2_{v,0.99} \leq \chi^2_{v,\alpha} \leq \chi^2_{v,0.05}\right]$

الفصل العاشر

مبادئ توزيعات المعاينة

ELEMENTARY OF SAMPLING DISTRIBUTIONS

١٠-١ مقدمة

تعد توزيعات المعاينة أساسا للاستدلال الاحصائي، لذا سنهتم هنا بعرض توزيعات المعاينة للعينات المختارة من مجتمع واحد فقط، بفرض أن لديناً مجتمعاً من المفردات يتبع توزيعا احتمالياً ما عدد عناصره N ونريد اختيار عينة حجمها n من هذا المجتمع، فإن عدد العينات التي يمكن اختيارها من هذا المجتمع هو:

- إذا كان الاختيار (السحب) بإرجاع (مع الإعادة) فإن عدد العينات يساوى N^n.
- إذا كان الاختيار (السحب) بدون إرجاع (بدون الإعادة) فإن عدد العينات يساوى $\binom{N}{n}$.

ومن المعلوم أن معالم المجتمع هي مقاييس يتميز بها المجتمع الإحصائي وهي ثابتة لنفس المجتمع لكن الإحصاءات هي مقاييس تتميز بها العينة وتختلف قيمة المقياس باختلاف العينة مع ثبوت حجمها لأنه يمكن اختيار أكثر من عينة من نفس المجتمع الإحصائي، لذا فإنه باختيار أكثر من عينة يكون لدينا مجتمعاً من الإحصاءات والتي هي بدورها تختلف في القيم عن بعضها البعض، ويمكن النظر إلى هذه المقاييس على أنها متغيرات عشوائية تتبع توزيعاً معينا قد يكون لها نفس التوزيع الاحتمالي للمجتمع الأصلي أو قد تختلف في التوزيع الاحتمالي عنه. ويسمى التوزيع الاحتمالي للإحصاءات بتوزيعات المعاينة. والإحصاء قد يكون الوسط الحسابي أو الانحراف المعياري للعينة أو نسبة مفردات العينة التي تتوفر فيها صفة معينة أو غيره من المقاييس الإحصائية.

مثال (١٠-١)

إذا كان لدينا مجتمعاً مكون من 20 عنصراً، تم اختيار عينات منه حجم كل منها 4 مفردات، فإذا تم تعيين الوسط الحسابي لكل عينة فأوجد حجم مجتمع المتوسطات؟

الحل

من الملاحظ أن لدينا البيانات التالية:

حجم المجتمع الأصلي $N=20$، حجم العينة $n=4$

وبالتالي فإن عدد العينات التي يمكن اختيارها بالحجم 4 من هذا المجتمع هي

أولاً في حالة السحب مع الإعادة يساوى

$$N^n = 20^4 = 160000$$

ثانياً في حالة السحب بدون الإعادة يساوى

$$\binom{N}{n} = \binom{20}{4} = 4845$$

وحيث إننا نحسب الوسط الحسابي لكل عينة فإن مجتمع المتوسطات يتكون من 160000 مشاهدة في حالة السحب مع الإعادة أو يتكون من 4845 مشاهدة في حالة السحب بدون إعادة. وبالمقارنة نجد أن عدد عناصر مجتمع المتوسطات أكبر بكثير من عدد مفردات المجتمع الأصلي.

يجب أن نفرق بين العينات المختارة من المجتمعات الكبيرة أو اللانهائية وبين العينات المختارة من مجتمعات محدودة ونلاحظ أنه عند اختيار العينات مع الإعادة فإن المفردة يكون لها فرصة الظهور أكثر من مرة (على الأقل مرة واحدة) فيمكن اعتبار المجتمع الإحصائي كبيراً أو لا نهائيا لكن في حالة السحب بدون إعادة فتكون للمفردة فرصة الظهور على الأكثر مرة واحدة فيمكن اعتبار المجتمع محدوداً وهذا لأنه يوجد اختلاف في خصائص توزيعات المعاينة للعينات المختارة من مجتمعات كبيرة عن تلك المختارة من مجتمعات محدودة وهذا ما سيتضح لاحقا خلال هذا الفصل.

١٠-٢ التوزيع الاحتمالي لمتوسط العينة (The Probability Distribution for the Sample Mean)

بفرض أن لدينا مجتمعاً وأننا نهتم بأحد المتغيرات فى هذا المجتمع والذى نرمز له بالرمز X ولتكن قيم هذا المتغير والتى تمثل مفردات المجتمع هى $X_1, X_2, X_3, \ldots$ وبفرض أننا اخترنا عينة حجمها n من هذا المجتمع ثم حسبنا وسطها الحسابي فكان $\bar{x}_1$ ثم اخترنا عينة ثانية لها نفس الحجم وحسبنا وسطها الحسابي فكان $\bar{x}_2$ وقمنا بتكرار اختيار العينات بنفس الحجم وتعيين الوسط الحسابي لكل عينة سوف نحصل على مجتمع جديد يسمى مجتمع المتوسطات للعينات التى حجمها n والتى يمكن اختيارها من المجتمع الأصلي وستكون مفردات هذا المجتمع هى

$\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots$ ومجتمع متوسطات العينات كأي مجتمع آخر له توزيع احتمالي ويتمتع بصفات وخواص التوزيعات الاحتمالية وبالطبع له متوسط وانحراف معياري خاص به. وخلال الأجزاء التالية من هذا الفصل سوف نستعين ببعض النظريات والتي لن نهتم بعرض إثباتها.

نظرية (١٠-١)

إذا كان لدينا مجتمع من المفردات له المتوسط μ والتباين σ^2 فإذا تم سحب جميع العينات بحجم n فإن متوسط مجتمع المتوسطات يرمز له بالرمز $\mu_{\bar{x}}$ وتباين مجتمع المتوسطات هو $\sigma_{\bar{x}}^2$ وتعطى كالتالي:

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}}^2 = \begin{cases} \dfrac{\sigma^2}{n}, & \text{إذا كان المجتمع الأصلى كبيراً} \\ \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right), & \text{إذا كان المجتمع الأصلى محدوداً} \end{cases}$$

ومن النظرية السابقة فإن متوسط مجتمع متوسطات العينات هو نفسه متوسط المجتمع الأصلي وانحرافه المعياري هو الانحراف المعياري للمجتمع الأصلي مضروباً فى عامل معين. وهذا محقق مهما كان التوزيع الاحتمالي للمجتمع الأصلي.

ومن الملاحظ أن تباين مجتمع المتوسطات أقل من تباين المجتمع الأصلي، مما يؤدى الى أن مجتمع متوسطات العينات أكثر تجانساً من المجتمع الاصلي أى أن مفرداته متجانسة وغير متشتته إذا ما قورنت بمفردات المجتمع الأصلي وهذا من أهم الأسباب التى تجعلنا نعتمد على مجتمع المتوسطات فى الاستدلال حول متوسط المجتمع الأصلي. فإذا كان الانحراف المعياري للمجتمع الأصلي هو $\sigma = 4$ وحجم العينة هو $n=5$ فإن الانحراف المعياري لمجتمع متوسطات العينات فى حالة سحب مفردات العينات بإرجاع هى

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{5}} = 1.789$$

وهذا يبين أن مجتمع متوسطات العينات أكثر تجانساً من المجتمع الأصلي.

مثال (١٠-٢)

إذا كانت 1, 2, 3, 4, 5, 6 عناصر المجتمع فالمطلوب تعيين قيمة μ, σ^2 لهذا المجتمع ثم إيجاد التوزيع العينى لـ $\bar{X}$ للعينات ذات الحجم $n=2$ والتحقق من العلاقات التالية:

(أ) فى حالة السحب بدون إرجاع فإن

$$\mu_{\bar{x}} = \mu, \quad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

(ب) فى حالة السحب بإرجاع فإن

$$\mu_{\bar{x}} = \mu, \qquad \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

الحل

بالنسبة للمجتمع الأصلي

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$= \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

X	$X-\mu$	$(X-\mu)^2$
1	−2.5	6.25
2	−1.5	2.25
3	−0.5	0.25
4	0.5	0.25
5	1.5	2.25
6	2.5	6.25
$\sum X = 21$		$\sum (X-\mu)^2 = 17.5$

$$\therefore \sigma^2 = \frac{17.5}{6} = 2.9167$$

$$\Rightarrow \sigma = \sqrt{\sigma^2} = \sqrt{2.917} = 1.708$$

أولاً السحب بدون اعادة:

عدد العينات الممكنة التي يمكن اختيارها بدون إعادة هى

$$k_1 = \binom{N}{n} = \binom{6}{2} = 15$$

توجد 15 عينة فى هذه الحالة. ويمكن سرد هذه العينات ومتوسطاتها كما فى الجدول التالي

الوسط الحسابي للعينات	العينات
1.5, 2, 2.5, 3, 3.5, 2.5, 3, 3.5, 4, 3.5, 4, 4.5, 4.5, 5, 5.5	(1, 2), (1,3), (1,4), (1,5), (1,6), (2,3), (2,4), (2,5), (2,6), (3,4), (3,5), (3,6), (4,5), (4,6), (5,6)

الوسط الحسابي لمجتمع المتوسطات يمكن تعيينه كما يلي:

$\bar{x}_r f_r$	عدد العينات	الوسط الحسابي للعينات
1.5	1	1.5
2.0	1	2.0
5.0	2	2.5
6.0	2	3.0
10.5	3	3.5
8.0	2	4.0
9.0	2	4.5
5.0	1	5.0
5.5	1	5.5
52.5	15	المجموع

$$\mu_{\bar{x}} = E(\bar{X}) = \frac{1}{k_1}\sum_{r=1}^{l}\bar{x}_r f_r = \frac{52.5}{15} = 3.5 = \mu$$

تباين مجتمع المتوسطات يمكن تعيينة كما يلي:

$$\sigma_{\bar{x}}^2 = \frac{1}{k_1}\sum_{r=1}^{l}(\bar{x}_r - \mu_{\bar{x}})^2 f_r$$

$(\bar{x}_r - \mu_{\bar{x}})^2 f_r$	عدد العينات	الوسط الحسابي للعينات
4.00	1	1.5
2.25	1	2.0
2.00	2	2.5
0.50	2	3.0
0.00	3	3.5
0.50	2	4.0
2.00	2	4.5
2.25	1	5.0
4.00	1	5.5
17.5	15	المجموع

$$\Rightarrow \sigma_{\bar{x}}^2 = \frac{17.5}{15} = 1.167$$

وبحساب المقدار

$$\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{2.917}{2}\left(\frac{6-2}{6-1}\right) = 1.167$$

وبمقارنة المقدارين معاً نجد أن

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

ثانياً السحب بارجاع:

عدد العينات الممكنة التي يمكن احتيارها من المجتمع بإرجاع هى

$$k_2 = N^n = 6^2 = 36$$

وبالتالي فإنه يوجد لدينا 36 عينة يمكن سردها وتعيين متوسط كلٍّ منها كما يلي:

العينات	الوسط الحسابى للعينات
(1,1), (1,2), (1,3), (1,4), (1,5), (1,6),	1, 1.5, 2, 2.5, 3, 3.5,
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),	1.5, 2, 2.5, 3, 3.5, 4,
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),	2, 2.5, 3, 3.5, 4, 4.5,
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),	2.5, 3, 3.5, 4, 4.5, 5,
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),	3, 3.5, 4, 4.5, 5, 5.5,
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)	3.5, 4, 4.5, 5, 5.5, 6

الوسط الحسابي لمجتمع المتوسطات يمكن تعيينه كالتالي:

$$\mu_{\bar{x}} = E(\bar{X}) = \frac{1}{k_2}\sum_{s=1}^{m} \bar{x}_s f_s$$

الوسط الحسابى للعينات	عدد العينات	$\bar{x}_s f_s$
1.0	1	1
1.5	2	3
2.0	3	6
2.5	4	10
3.0	5	15
3.5	6	21
4.0	5	20
4.5	4	18
5.0	3	15
5.5	2	11
6.0	1	6
المجموع	36	126

$$\Longrightarrow \mu_{\bar{x}} = \frac{126}{36} = 3.5 = \mu$$

تباين مجتمع المتوسطات يمكن تعيينة كالتالي:

$$\sigma_{\bar{x}}^2 = Var(\bar{X}) = \frac{1}{k_2}\sum_{s=1}^{m}(\bar{x_s} - \mu_{\bar{x}})^2 f_s$$

الوسط الحسابي للعينات	عدد العينات	$(\bar{x_s} - \mu_{\bar{x}})^2 f_s$
1.0	1	6.25
1.5	2	8.00
2.0	3	6.75
2.5	4	4.00
3.0	5	1.25
3.5	6	0.00
4.0	5	1.25
4.5	4	4.00
5.0	3	6.75
5.5	2	8.00
6.0	1	6.25
المجموع	36	52.5

$$\Longrightarrow \sigma_{\bar{x}}^2 = \frac{52.5}{36} = 1.4583$$

وبحساب المقدار

$$\frac{\sigma^2}{n} = \frac{2.9167}{2} = 1.4583$$

وبمقارنة المقدارين معاً نجد أن:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

حتى الآن كل ما توصلنا له فقط هو المتوسط والتباين لمجتمع المتوسطات والتوزيع الاحتمالي لمجتمع متوسطات العينات يعتمد بشكل أساس على توزيع المجتمع الأصلي التى اختيرت منه العينات. والنظرية التالية توضح ذلك.

نظرية (١٠-٢)

إذا كان لدينا مجتمع يتبع التوزيع الطبيعي بمتوسط μ وتباين مقداره σ^2 وتم اختيار عينات عشوائية حجم كلّ منها n من هذا المجتمع فإن الوسط الحسابي للعينات يتبع أيضاً التوزيع الطبيعي بمتوسط $\mu_{\bar{x}}$ وتباين مقداره $\sigma_{\bar{x}}^2$.

ومن النظرية السابقة يتضح أنه إذا كان $X \sim N(\mu, \sigma^2)$ فإن $\bar{X} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$ فإن

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \sim N(0,1)$$

مثال (١٠-٣)

إذا كانت أطوال طلاب الجامعة تتبع التوزيع الطبيعي بمتوسط 170 سنتميترا وانحراف معياري مقداره 5 سنتيمترات. فاخترنا عينة مكونة من 25 طالباً، فما هو احتمال أن يكون متوسط أطوالها أكبر من 171.5سنتيمتراً؟

الحل

بفرض أن X ترمز إلى أطوال طلاب الجامعة فإن $X \sim N(\mu, \sigma^2)$ بحيث $\mu = 170, \sigma = 5, n = 25$ فإن $\bar{X} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$ وحيث إن حجم المجتمع غير محدود فإن

$$\mu_{\bar{x}} = \mu = 170,$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{(5)^2}{25} = 1$$

والمطلوب تعيين قيمة$P[\bar{X} < 180]$ والتي يمكن تعيينها كما يلي:

$$P[\bar{X} > 180] = P\left[\frac{\bar{X} - \mu}{\sigma_{\bar{x}}} > \frac{171.5 - 170}{1}\right]$$

$$= P[Z > 1.5] = 1 - \Phi(1.5) = 0.0668$$

لقد تعرضنا فى النظرية السابقة للتوزيع الاحتمالى للوسط الحسابى $\bar{X}$ إذا كانت العينة مختارة من مجتمع يتبع توزيعاً طبيعياً، ولكن فى كثير من الحالات يكون التوزيع الاحتمالى للمجتمع الأصلى مجهولاً ويتطلب الأمر معرفة التوزيع الاحتمالى للوسط الحسابى ففى مثل هذه الحالات نستخدم نظرية تعتبر من النظريات الهامة فى علم الاحصاء وتسمى بنظرية النهاية المركزية.

نظرية(١٠-٣) نظرية النهاية المركزية (Central Limit Theorem)

إذا كان لدينا مجتمع يتبع توزيعاً احتمالياً متوسطه μ وتباينه σ^2 واخترنا منه عينات عشوائية حجم كلاً منها n وكانت n كبيرة. فإن الوسط الحسابي $\bar{X}$ لهذه العينات يتبع تقريباً توزيعاً طبيعياً وسطه $\mu_{\bar{x}}$ وتباينه $\sigma_{\bar{x}}^2$.

أى أن $\bar{X}\sim N(\mu_{\bar{x}},\sigma_{\bar{x}}^2)$ عندما $n\longrightarrow\infty$ فإن المتغير العشوائى

$$Z=\frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}^2}\sim N(0,1)$$

ومن المعلوم أنه إذا كانت $n\geq 30$ فهذا يعنى أن $n\longrightarrow\infty$.

مثال(١٠-٤)

إذا كان متوسط الدخل الشهرى للأسر في إحدى المدن 600 دولار والانحراف المعياري هو 100 دولار اختيرت عينة عشوائية حجمها 100 أسرة من هذه المدينة، فأوجد احتمال أن يتراوح متوسط دخل الأسرة فى العينة بين 595 ,615 دولاراً.

الحل

بفرض أن X يرمز للدخل الشهرى للأسرة فإن

$$n=100,\mu=600,\sigma=100$$

والمطلوب تعيين $P[595<\bar{X}<615]$

وحيث إن المجتمع غير طبيعي، لكن n>30 فإنه باستخدام نظرية النهاية المركزية فإن $\bar{X}$ يتوزع توزيعاً طبيعياً وسطه $\mu_{\bar{x}}$ وتباينه $\sigma_{\bar{x}}^2$ ومن ثم فإن $\bar{X}\sim N(\mu_{\bar{x}},\sigma_{\bar{x}}^2)$ وحيث إن حجم المجتمع غير محدود فإن

$$\mu_{\bar{x}}=\mu=600,$$

$$\sigma_{\bar{x}}^2=\frac{\sigma^2}{n}=\frac{(100)^2}{100}=100\implies\sigma_{\bar{x}}=\sqrt{100}=10$$

$$\therefore P[595<\bar{X}<615]=P\left[\frac{595-600}{10}<\frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}<\frac{615-600}{10}\right]$$

$$=P[-0.5<Z<1.5]=\Phi(1.5)-\Phi(-0.5)$$

$$=\Phi(1.5)-[1-\Phi(0.5)]$$

$$=0.9332-1+0.6915=0.6247$$

مثال(١٠-٥)

إذا كان ضغط الدم لمجموعة من 1000 فرد متوسطه 90 وانحرافه المعياري 9 واخترنا عينة حجمها 36 فرداً من هذه المجموعة. أوجد احتمال أن الوسط الحسابي لضغط الدم فى العينة أكبر من 86؟

الحل

بفرض أن X يرمز للدخل الشهري للأسرة فإن $n = 36, \mu = 90, \sigma = 9,\ N = 1000$ والمطلوب تعيين $P[\bar{X} > 86]$ وحيث إن المجتمع غير طبيعي، n>30 فإنه باستخدام نظرية النهاية المركزية فإن $\bar{X}$ يتوزع توزيعاً طبيعياً وسطه $\mu_{\bar{x}}$ وتباينه $\sigma_{\bar{x}}^2$ ومن ثم فإن $\bar{X} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$ وحيث إن حجم المجتمع محدود فإن

$$\mu_{\bar{x}} = \mu = 90,$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{(9)^2}{36}\left(\frac{1000-36}{1000-1}\right) = 2.1712$$

$$\Longrightarrow \sigma_{\bar{x}} = \sqrt{2.1712} = 1.4735$$

$$\therefore P[\bar{X} > 86] = P\left[\frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{86-90}{1.4735}\right]$$

$$= P[Z > -2.7146] = 1 - \Phi(-2.7146)$$

$$= 1 - [1 - \Phi(2.7146)] = \Phi(2.7146) = 0.9966$$

مثال(١٠-٦)

إذا كانت الحوادث الأسبوعية على أحد الطرق تتبع توزيع بواسون بمتوسط أربعة حوادث وأخذت عينة من الحوادث فى 36 اسبوعاً. فما احتمال أن يكون متوسط عدد الحوادث يزيد عن 3.2 حادثاً؟

الحل

بفرض أن X يمثل عدد الحوادث الأسبوعية على أحد الطرق فإن X يتبع توزيع بواسون بمتوسط 4 حوادث. فمن خواص توزيع بواسون

$$\mu = \sigma^2 = 4$$

وحيث إن n= 36 >30 فمن نظرية النهاية المركزية فإن $\bar{X} \sim N(\mu_{\bar{x}}, \sigma_{\bar{x}}^2)$ وحيث إن المجتمع غير محدود فإن

$$\mu_{\bar{x}} = \mu = 4,$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{4}{36} = \frac{1}{9} \Longrightarrow \sigma_{\bar{x}} = \sqrt{\frac{1}{9}} = \frac{1}{3} = 0.3333$$

$$\therefore P[\bar{X} \geq 3.2] = P\left[\frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} \geq \frac{3.2-4}{0.3333}\right]$$

$$= P[Z \geq -2.4] = 1 - \Phi(-2.4)$$

$$= 1 - [1 - \Phi(2.4)] = \Phi(2.4) = 0.9918$$

من الملاحظ أن المقدار $Z = \frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}$ يحتوى على متغير واحد فقط وهو $\bar{X}$ والذى يختلف من عينة لأخرى وكل من μ, σ^2 معلومتان. لكن إذا كانت σ^2 مجهولة (كما هو الحال فى أغلب التطبيقات العملية) فإنه يجب تقديرها من مشاهدات العينة باستخدام تباين العينة s^2 والتى تعطى بالعلاقة التالية:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

فإذا كانت العينة كبيرة أى $n \geq 30$ فإن s^2 تكون قريبة جداً من σ^2 وفى هذه الحالة يكون المقدار

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{s_{\bar{x}}} \sim N(0,1)$$

بحيث

$$\mu_{\bar{x}} = \mu$$

$$s_{\bar{x}}^2 = \begin{cases} \frac{s^2}{n}, & \text{إذا كان المجتمع الأصلي كبيراً} \\ \frac{s^2}{n}\left(\frac{N-n}{N-1}\right), & \text{إذا كان المجتمع الأصلي محدوداً} \end{cases}$$

مثال(١٠-٧)

إذا كان متوسط درجات الطلاب فى مقرر 1080 إحص هو 80 درجة، تم اختيار الشعبة رقم 200 من الطلاب الذين يدرسون هذا المقرر وكانت تحتوى على 34 طالباً وكان الانحراف المعياري لدرجاتهم هو 4 درجات فأوجد احتمال أن تقل درجاتهم عن 81 درجة.

الحل

بفرض أن X يرمز لدرجات الطلاب فى مقرر 1080 إحص فإن $n = 34, \mu = 80, s = 4,$ والمطلوب تعيين قيمة $P[\bar{X} < 81]$ وحيث إن المجتمع غير طبيعي، والتباين مجهول، لكن $n>30$ فإنه باستخدام نظرية النهاية المركزية فإن $\bar{X}$ يتوزع توزيعاً طبيعياً وسطه $\mu_{\bar{x}}$ وتباينه $s_{\bar{x}}^2$ ومن ثم فإن $\bar{X} \sim N(\mu_{\bar{x}}, s_{\bar{x}}^2)$ وحيث إن حجم المجتمع غير محدود فإن

$$\mu_{\bar{x}} = \mu = 80,$$

$$s_{\bar{x}}^2 = \frac{s^2}{n} = \frac{(4)^2}{34} = 0.4706$$

$$\Longrightarrow s_{\bar{x}} = \sqrt{0.4706} = 0.6859$$

$$\therefore P[\bar{X} < 81] = P\left[\frac{\bar{X} - \mu_{\bar{x}}}{s_{\bar{x}}} < \frac{81 - 80}{0.6859}\right]$$

$$= P[Z < 1.4579] = \Phi(1.46) = 0.9279$$

لقد توصلنا سابقاً للتوزيع الاحتمالي لمجتمع متوسطات العينات إذا كان المجتمع الأصلي يتبع التوزيع الطبيعي بتباين معلوم وأيضاً باستخدام نظرية النهاية المركزية حصلنا على التوزيع الاحتمالي لمجتمع متوسطات العينات إذا كان حجم العينات كبيرة بغض النظر عن التوزيع الاحتمالي للمجتمع الأصلي طبيعياً أو غير طبيعي، لكنه إذا كان حجم العينات صغيراً فلا يمكن الاستعانة بنظرية النهاية المركزية ولكننا سوف نستخدم النظرية التالية.

نظرية(١٠-٤)

إذا كان $x_1, x_2, \dots, x_n$ عينة عشوائية تم اختيارها من مجتمع يتبع التوزيع الطبيعي متوسطة μ وتباينة σ^2 وكانت σ^2 مجهولة وحجم العينة n صغيراً فإن المتغير:

$$\frac{\bar{X} - \mu_{\bar{x}}}{s_{\bar{x}}}$$

يتبع توزيع t بدرجة حرية $n-1$. أى ان $X \sim N(\mu, \sigma^2)$ فإن $\frac{\bar{X}-\mu_{\bar{x}}}{s_{\bar{x}}} \sim t_{n-1}$ وذلك إذا كانت σ^2 مجهولة، n صغيرة $(n < 30)$.

مثال(١٠-٨)

إذا علمت أن علامات الطلاب فى مقرر 1040 إحص تخضع للتوزيع الطبيعي بمتوسط 70 أخذت عينة حجمها 16 طالباً فوجد أن الانحراف المعياري لعلاماتهم هى 8، احسب احتمال أن يزيد متوسط علاماتهم عن 74.

الحل

بفرض أن X متغير عشوائى يمثل درجات الطلاب فى مقرر 1040 إحص بحيث $X \sim N(\mu, \sigma^2)$ حيث $\mu = 70, n = 16, s = 8$ والمطلوب هو تعيين $P[\bar{X} > 74]$ وحيث إن المجتمع الأصلى يتبع التوزيع الطبيعي وتباينه مجهول وأيضاً $n < 30$ فإن

$$\frac{\bar{X} - \mu_{\bar{x}}}{s_{\bar{x}}} \sim t_{n-1}$$

حيث

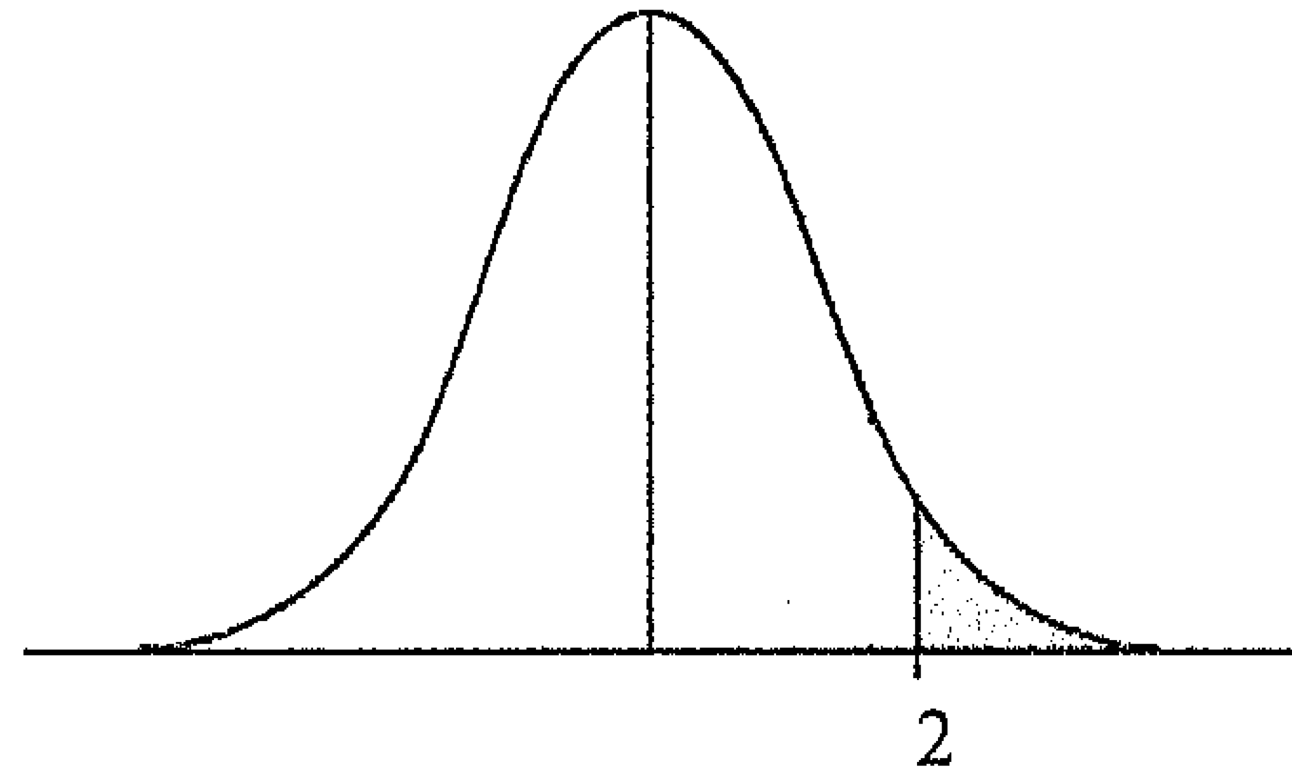
$$\mu_{\bar{x}} = \mu = 70,$$

$$s_{\bar{x}}^2 = \frac{s^2}{n} = \frac{(8)^2}{16} = 4 \Longrightarrow s_{\bar{x}} = \sqrt{4} = 2$$

$$\therefore P[\bar{X} > 74] = P\left[\frac{\bar{X} - \mu_{\bar{x}}}{s_{\bar{x}}} > \frac{74 - 70}{2}\right]$$

$$= P[Y > 2] = \alpha$$

من منحنى توزيع t فإن



حيث

$$2 = t(n - 1, \alpha) = t(15, \alpha)$$

من جدول توزيع t نجد أن $\alpha = 0.025$ فإن:

$$P[\bar{X} > 74] = 0.025$$

مثال(١٠-٩)

إذا كان X متغيراً عشوائياً يمثل أطوال طلاب المرحلة المتوسطة وكان X يتبع التوزيع الطبيعي بمتوسط 150 سنتيمتراً. أخذت عينة حجمها 9 طلاب من إحدى مدارس المرحلة المتوسطة، ووجد أن الانحراف المعياري لها يساوى 5 سنتيمتراً. احسب احتمال أن يزيد الوسط الحسابى لهذه العينة عن 155سنتيمتراً.

الحل

من المعطيات فإن $\mu = 150, n = 9, s = 5$ والمطلوب حساب $P[\bar{X} \geq 155]$

بما أن المجتمع يتبع التوزيع الطبيعي معلوم متوسطه لكن تباينة مجهول وأيضاً n< 30 فإن

$$\frac{\bar{X} - \mu_{\bar{x}}}{s_{\bar{x}}} \sim t_{n-1}$$

حيث

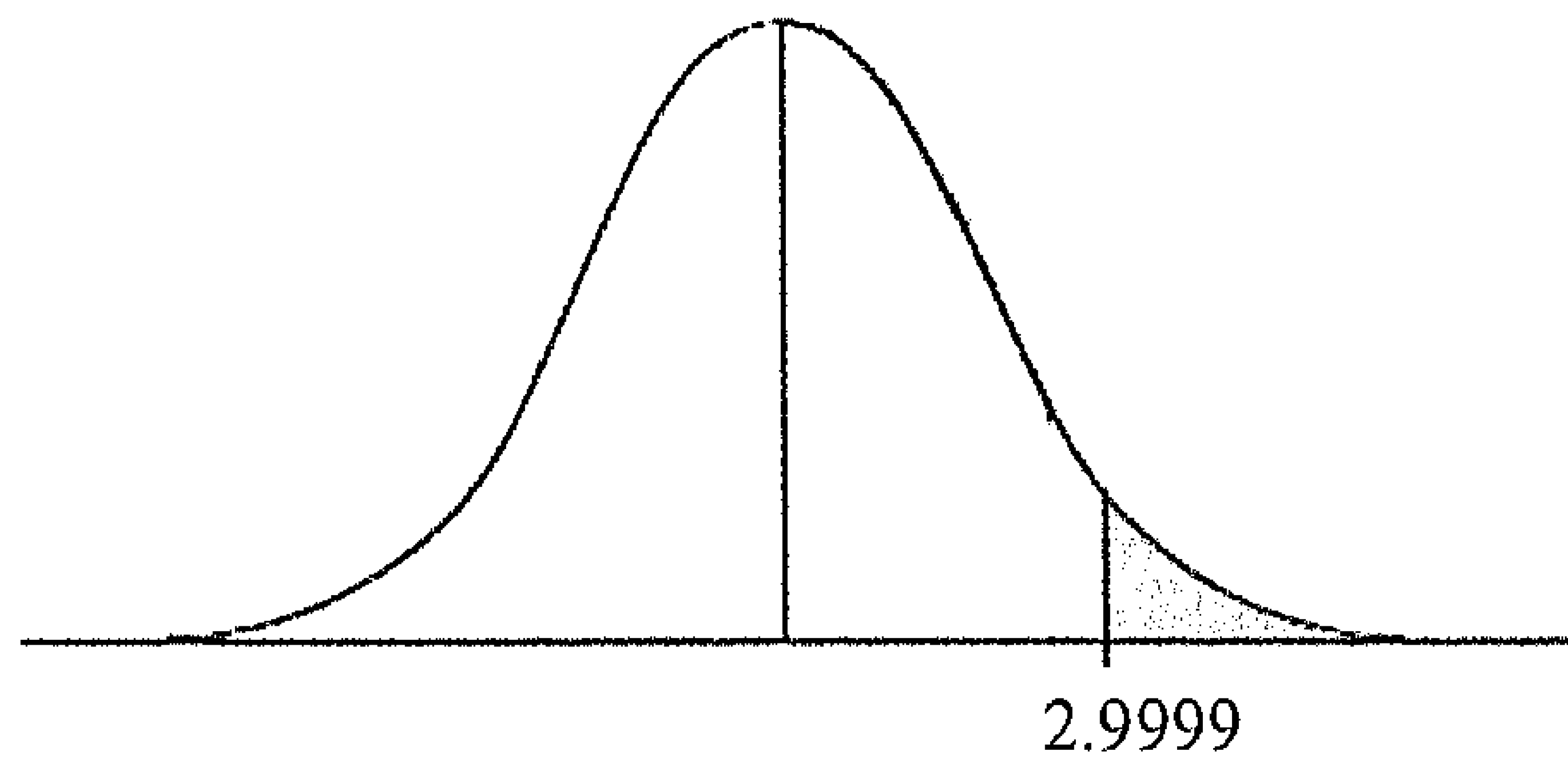
$$\mu_{\bar{x}} = \mu = 150,$$

$$s_{\bar{x}}^2 = \frac{s^2}{n} = \frac{(5)^2}{9} = 2.7778 \Longrightarrow s_{\bar{x}} = \sqrt{2.7778} = 1.6667$$

$$\therefore P[\bar{X} \geq 155] = P\left[\frac{\bar{X} - \mu_{\bar{x}}}{s_{\bar{x}}} \geq \frac{155 - 150}{1.6667}\right]$$

$$= P[Y \geq 2.9999] = \alpha$$

من منحنى توزيع t فإن



$$2.9999 = t(\alpha, n-1) = t(\alpha, 8)$$

ومن جدول توزيع t فإن $\alpha = 0.01$ وبالتالي فإن

$$P[\bar{X} \geq 155] = 0.01$$

١٠-٣ التوزيع الاحتمالي للنسبة في العينة (Sampling Distribution of the Sample Proportion)

بفرض أن لدينا مجتمعاً ما وأن بعض مفرداته تتوفر فيها صفة معينة وأن نسبة هذه المفردات هي p، كأن تكون p نسبة الأفراد الأميين فى إحدي المدن، أو نسبة الأجهزة التالفة التى ينتجها مصنع ما. فإذا اخذنا عينة عشوائية حجمها n من هذا المجتمع ووجدنا من بينها x مفردة تتوفر فيها هذه الصفة فإن النسبة $r = \frac{x}{n}$ تكون هى نسبة مفردات العينة التى تتوفر فيها الصفة المعينة. فإذا أخذنا كل العينات الممكنة والمتساوية الحجم n من المجتمع

الذى حجمه N وبتعيين نسبة تلك الصفة فى كل عينة ولتكن $r_1, r_2, \ldots$ فإن هذه النسبة سوف تختلف من عينة لأخرى فيصبح لدينا متغير عشوائي R للنسبة يأخذ القيم المختلفة للعينات $r_1, r_2, \ldots$ بحيث

$$\mu_r = E(R) = p,$$

$$\sigma_r^2 = Var(R) = \begin{cases} \dfrac{p(1-p)}{n}, & \text{إذا كان المجتمع غير محدود} \\ \dfrac{p(1-p)}{n}\left(\dfrac{N-n}{N}\right), & \text{إذا كان المجتمع محدوداً} \end{cases}$$

نظرية(١٠-٥)

إذا كانت p هى نسبة وجود ظاهرة معينة فى مجتمع ما واختير من هذا المجتمع عينات كبيرة حجم كل منها n وكانت r تمثل نسبة وجود هذه الظاهرة فى العينات فإن r تتبع تقريباً التوزيع الطبيعي بمتوسط μ_r وتباين σ_r^2.

ومن النظرية السابقة فإن $r \sim N(\mu_r, \sigma_r^2)$ وبالتالي فإن:

$$Z = \frac{r - \mu_r}{\sigma_r} \sim N(0,1)$$

مثال(١٠-١٠)

إذا علمت أن نسبة الوحدات المعيبة لإنتاج إحدي المؤسسات هو 0.12، واشترى شخص 40 وحدة من هذه المؤسسة، فما احتمال أن يجد من بينها 16 وحدة معيبة على الأكثر؟

الحل

من المعطيات a= 16, n = 40, p =0.12 فإن النسبة فى العينة هي

$$r = \frac{a}{n} = \frac{16}{100} = 0.16$$

والمطلوب تعيين قيمة

$$P[a \leq 16] = P[r \leq 0.16]$$

وحيث إن حجم العينة كبير فإن $r \sim N(\mu_r, \sigma_r^2)$ بحيث

$$\mu_r = p = 0.12,$$

$$\sigma_r^2 = \frac{p(1-p)}{n} = \frac{0.12(1-0.12)}{40} = 0.00264$$

$$\Longrightarrow \sigma_r = \sqrt{0.00264} = 0.05138$$

$$\therefore P[r \le 0.16] = P\left[\frac{r-\mu_r}{\sigma_r} \le \frac{0.16-0.12}{0.05138}\right]$$

$$= P[Z \le 0.7785]$$

$$= \Phi(0.78) = 0.823$$

مثال(١٠-١١)

إذا كان احتمال رسوب الطالب الذى يدرس مقرر 1080 إحص هو 0.30 وأخذت عينة حجمها 36 طالباً من أولئك الذين يدرسون هذا المقرر فأوجد احتمال أن تقل نسبة الراسبين فى هذه العينة عن 0.4.

الحل

بفرض أن نسبة الراسبين فى مقرر 1080 إحص هى p حيث $p = 0.30, n = 36$ والمطلوب تعيين قيمة $P[r \le 0.4]$ وحيث إن حجم العينة كبير فإن $r \sim N(\mu_r, \sigma_r^2)$ بحيث

$$\mu_r = p = 0.30,$$

$$\sigma_r^2 = \frac{p(1-p)}{n} = \frac{0.3(1-0.3)}{36} = 0.00583$$

$$\Longrightarrow \sigma_r = \sqrt{0.00583} = 0.0764$$

$$\therefore P[r \le 0.4] = P\left[\frac{r-\mu_r}{\sigma_r} \le \frac{0.4-0.3}{0.0764}\right]$$

$$= P[Z \le 1.3089]$$

$$= \Phi(1.31) = 0.9049$$

١٠-٤ التوزيع الاحتمالي لتباين العينة (Sampling Distribution of the Sample Variance)

يستخدم التباين لدراسة تجانس البيانات، ويقوم بتعيين تشتت القيم بالنسبة لوسطها الحسابى. وبفرض أن لدينا مجتمعاً ما متوسطه μ وتباينه σ^2 فإذا أخذنا كل العينات الممكنة والمتساوية الحجم n من المجتمع الذى حجمه N وبتعيين التباين لكل عينة ولتكن $s_1^2, s_2^2, \dots$ فإن هذه التباينات سوف تختلف من عينة لأخرى فيصبح

لدينا متغير عشوائي S^2 للتباين فنحصل على مجتمع جديد يسمى مجتمع التباينات والذى يمكن تعيين التوزيع الاحتمالي له باستخدام النظرية التالية:

نظرية(١٠-٦)

إذا كانت ... $x_1, x_2,$ عينة عشوائية تم اختيارها من مجتمع له التوزيع الطبيعي بمتوسط μ وتباين σ^2 فإن المتغير $\frac{(n-1)s^2}{\sigma^2}$ يتبع توزيع (مربع كاي) بدرجة حرية $(n-1)$.

مثال(١٠-١٢)

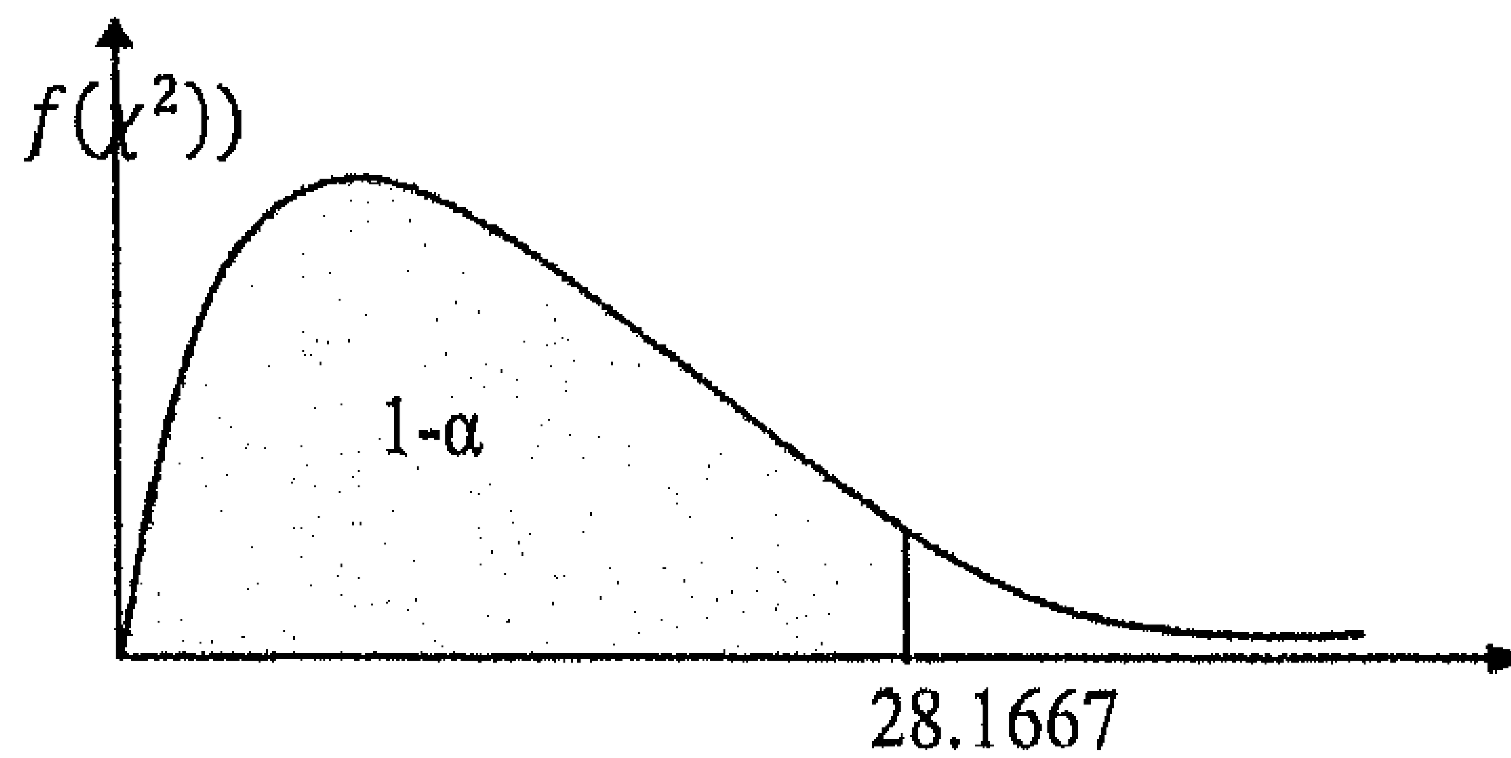
ينتج مصنع للأدوية نوعاً من العلاج يحتوى على مادة فعالة ويجب أن تكون كمية هذه المادة محددة بشكل دقيق. ولدراسة مدى دقة هذا المصنع فى إضافة هذه الكمية إلى كل حبة من حبوب هذا العلاج قام مدير المصنع بتحليل عينة من 25 حبة. فاحسب احتمال أن لا يزيد الانحراف المعيارى لكمية هذه المادة فى العينة عن 1.30 مليجراماً علماً بأن وزن هذه المادة فى إنتاج المصنع كافة يتبع التوزيع الطبيعي بانحراف معياري يساوى 1.20 مليجراماً.

الحل

بفرض أن X متغير عشوائى يمثل وزن المادة الفعال فى إنتاج المصنع كافة فإن $X \sim N(\mu, \sigma^2)$ بحيث $\sigma = 1.2,\ n = 25$ والمطلوب تعيين $P[s \leq 1.3]$

$$\therefore P[s \leq 1.3] = P[s^2 \leq 1.69]$$
$$= P\left[\frac{(n-1)s^2}{\sigma^2} \leq \frac{(25-1)(1.69)}{(1.2)^2}\right]$$
$$= P[\chi^2 \leq 28.1667]$$

من منحنى توزيع مربع كاى فإن



$$P[s \leq 1.3] = P[\chi^2 \leq 28.1667] = 1 - \alpha$$

حيث

$$28.1667 = \chi^2(n-1, \alpha) = \chi^2(24, \alpha) \Rightarrow \alpha = 0.05$$

وبالتالى فإن

$$P[s \leq 1.3] = 1 - \alpha = 1 - 0.05 = 0.95$$

أسئلة وتمارين (١٠)

١- أيهما أكثر تجانساً: المجتمع الأصلى أم مجتمع الأوساط الحسابية للعينات؟ ولماذا؟

٢- ما الشرط الضرورى لصحة نظرية النهاية المركزية؟

٣- إذا كان لدينا مجتمع متوسطه 100 وتباينه 4 أخذت منه جميع العينات بالحجم 3 فأوجد متوسط وتباين مجتمع المتوسطات.

٤- أخذت عينة حجمها 2 من المجتمع الذى عناصره {1, 2, 3, 4} فإذا تم اختيار العينة بإرجاع فأوجد

(أ) جميع العينات الممكنة.

(ب) التوقع الرياضي والتباين لمجتمع المتوسطات.

٥- نسبة الذكاء فى مجتمع ما يكون لها توزيع طبيعي بمتوسط 100 وانحراف معياري 10 فإذا اخترنا عينة حجمها 16 فرداً من هذا المجتمع فما احتمال أن يقل متوسط الذكاء لهذه العينة عن 104 درجة؟

٦- إذا كانت علامات الطلاب فى مقرر 1080 إحص تخضع لتوزيع طبيعي وسطه 80 وأخذت عينة حجمها 16 طالباً فوجد أن الانحراف المعياري لعلاماتهم 8 احسب احتمال أن يزيد الوسط الحسابي للعينة عن 74.

٧- إذا كان X متغيراً عشوائياً وسطه 185.6 وانحرافه المعياري 12.7 فما احتمال أن الوسط الحسابى لعينة حجمها 36 أخذت من هذا التوزيع يزيد عن 185.

٨- إذا كان X متغيراً عشوائياً يمثل أوزان أكياس الأرز التى تنتجها إحدى المؤسسات، وكان X خاضعاً للتوزيع الطبيعي بمتوسط 50 كيلوجراماً وأخذت عينة حجمها 9 أكياس من إنتاج هذه المؤسسه ووجد أن الانحراف المعياري لأوزان هذه الأكياس يساوى 2 كيلوجرام. احسب احتمال أن يزيد الوسط الحسابي لهذه العينة عن 52 كيلوجراماً.

٩- إذا كانت $x_1, x_2, \ldots, x_{10}$ عينة عشوائية من توزيع طبيعي $N(\mu, 9)$ تباينها هو s^2 فأوجد قيمة الثابت c إذا كانت $P[s^2 \leq c] = 0.99$.

١٠- إذا كان X متغيراً عشوائياً يتبع التوزيع الطبيعي بمتوسط 160 وانحراف معياري 12 فما احتمال أن الوسط الحسابي لعينة حجمها 14 أخذت من هذا التوزيع يزيد عن 165.

١١- إذا كان أطوال 200 من الطلاب تخضع لتوزيع طبيعي وسطه 160 سنتيمتراً تم اختيار عينة من الطلاب حجمها 40 طالباً فكان الانحراف المعياري لهم هو 4 فأوجد احتمال أن يزيد الوسط الحسابي لأطوال العينة عن 165 سنتيمتراً.

١٢- إذا كانت درجات طلاب كلية العلوم والدراسات الإنسانية بجامعة الخرج فى مقرر 1080 كيم تخضع لتوزيع طبيعي وسطه 70 وبتاين مقداره 16 وأخذت عينة حجمها 16 طالباً ممن يدرسون هذا المقرر. أوجد احتمال أن تكون درجاتهم أكبر من 72.

١٣- إذا كانت اجور المعلمين فى وزارة التربية والتعليم تخضع للتوزيع الطبيعي بمتوسط 2100 ريالاً سعودياً. أخذت عينة حجمها 25 معلماً من معلمى الوزارة فوجد أن الانحراف المعياري لها هو 400 ريالاً سعودياً أوجد احتمال أن يكون متوسط أجورهم يزيد عن 2300 ريالاً.

١٤- إذا كانت نسبة المدخنين فى المملكة العربية السعودية هى 0.3 أخذت عينة من سكان الرياض حجمها 36 فرداً فأوجد احتمال أن يزيد عدد المدخنين فى تلك العينة عن 14 شخصاً.

١٥- إذا كان أطوال الطلاب في إحدى الجامعات تخضع لتوزيع طبيعي متوسطه 160 سنتيمتراً وانحراف معياري مقداره 10سنتيمتراً وأخذت عينة حجمها 16 طالباً من طلاب هذه الجامعة. فأوجد احتمال أن:

(أ) المتوسط الحسابي لأطوال طلاب العينة أكبر من 156 سنتيمتراً.

(ب) تباين العينة يزيد عن 12 سنتيمتراً.

١٦- إذا كانت $x_1, x_2, \dots, x_{20}$ عينة عشوائية من مجتمع يتبع التوزيع الطبيعي بتباين مقداره 1 فأوجد قيمة الثابت c بحيث أن $P[s^2 \leq c] = 0.9$.

١٧- أخذت عينة عشوائية حجمها 25 طفلاً ممن هم فى عامهم الأول، فإذا كانت أوزان الأطفال فى هذا السن تخضع لتوزيع طبيعي انحرافه المعياري 500 جرام. فأوجد احتمال أن يقل الانحراف المعيارى لأوزان العينة عن 400 جرام.

١٨- إذا كانت نسبة طلاب الحاصلين على معدل +A فى مقرر الإحصاء 1040 إحص فى كلية العلوم هى 0.12 أخذت عينة من الذين يدرسون هذا المقرر حجمها 36 طالباً فأوجد احتمال أن يزيد عدد الحاصلين على معدل +A فى تلك العينة عن 14 طالباً.

١٩- إذا كانت نسبة المصابين بمرض ما فى المملكة العربية السعودية هى 0.2 أخذت عينة من سكان مدينة الرياض حجمها 50 فرداً. فأوجد احتمال أن يقل عدد المصابين بهذا المرض فى تلك العينة عن 19 شخصاً.

٢٠- أخذت عينة عشوائية حجمها 45 طالباً ممن هم فى المستوي الثاني، فإذا كانت درجاتهم فى الاختبار النهائي لمقرر 1080 إحص تخضع لتوزيع طبيعي انحرافه المعياري 5 درجات فأوجد احتمال أن يقل الانحراف المعياري لدرجاتهم فى العينة عن 6 درجات.

٢١- إذا كانت أطوال طلاب الجامعة تخضع لتوزيع طبيعي وسطه 170 سنتيمتراً تم اختيار عينة من الطلاب حجمها 20 طالباً فكان الانحراف المعياري لهم هو 4 سم. فأوجد احتمال أن يقل الوسط الحسابي لأطوال العينة عن 165 سنتيمتراً.

الفصل الحادي عشر

مبادئ الاستدلال الإحصائي

ELEMENTARY STATISTICAL INFERENCE

١١-١ مقدمة

يرتكز البحث العلمي فى العديد من مجالاته على الطرق الإحصائية كأدوات لا غني عنها في استخلاص المعلومات. وقد زادت الحاجة لهذه الطرق بصفة خاصة مع تزايد الاعتماد على المعاينة كأسلوب فى جمع البيانات بدلاً من أسلوب الحصر الشامل. ذلك لأن محاولة تعميم النتائج الناتجة من العينة على المجتمع لا يمكن أن تخلو من الأخطاء، وهي تتطلب استحداث وسائل تتيح إجراء عملية التعميم بأقل قدر ممكن من المخاطرة. وإذا كان الإحساس النظري الذى تستند إليه هذه الطرق وهو ما يسمى بالنظرية الإحصائية يتم التوصل إليه في إطار ما يعرف بالتبرير الاستنباطي (deductive reasoning) والذي تستنبط فيه الحقائق كنتائج منطقية لمسلمات، فإن استخدامها فى تعميم نتائج العينة يسمى بالتبرير الاستقرائي (inductive reasoning) ويطلق على عملية تعميم نتائج العينة للمجتمع بالاستدلال الإحصائي (الإستقراء- statistical inference). وينقسم الاستدلال الإحصائي إلي جزئين أساسين هما: التقدير (estimation) واختبارات الفروض (hypotheses tests)؛ فأحيانا يكون لدينا مجتمع غير معلوم معالمه الإحصائية ونريد تقدير تلك المعالم لإجراء دراسة معينة وهذا ما يسمى بالتقدير، وأيضاً قد يكون لدينا ادعاء حول معالم المجتمع الإحصائي كأن يقول شخص ما بأن نسبة المصابين بمرض معين هي 30% ونريد اختبار هذا الادعاء وهذا ما يطلق عليه اختبارات الفروض الإحصائية. وسوف نتعرض خلال هذا الفصل لدراسة التقدير واختبارات الفروض الإحصائية حول معالم مجتمع إحصائي واحد فقط.

١١-٢ مبادئ نظرية التقدير (Elementary of Esimation Theory)

عادة ما تكون بعض معالم المجتمع الإحصائي مجهولة، فقد يكون الوسط الحسابى μ أو التباين σ^2 أو نسبة ظهور صفة ما p فى المجتمع مجهولة مثل متوسط عمر المصباح الذى ينتجة مصنع معين، أو متوسط الدخل فى مجتمع ما، أو نسبة المصابين بمرض معين، أو غير ذلك. لذا فإن هدف الإحصائي معرفة المعالم المجهولة ولكن ليس من اليسير معرفة قيم هذه المعالم بالضبط لهذا نضطر إلي تقدير تلك القيم. ويمكن تقدير قيمة المعلمة المجهولة بإحدى طريقتين: إما التقدير بنقطة وهي أن نختار قيمة واحدة فقط كتقدير لمعلمة المجتمع المجهولة، فيمكن استخدام متوسط عمر حشرة من نوع معين والمحسوب من عينة عشوائية مسحوبة من مجتمع هذا النوع من الحشرات، كتقدير لمتوسط عمر الحشرات من هذا النوع وإننا سوف نستخدم الإحصاءات المقابلة للعينة لتقدير معالم المجتمع المجهول. فسوف نستخدم متوسط العينة $\bar{x}$ لتقدير متوسط المجتمع μ وأيضاً نستخدم تباين العينة لتقدير تباين المجتمع المجهول ونسبة صفة ما فى العينة لتقدير نفس الصفة فى المجتمع المسحوب منه العينة وهكذا. وتكتب $\hat{\mu} = \bar{X}$ ، $\widehat{\sigma^2} = s^2$ وأيضاً $\hat{p} = r$ وتختلف الإحصاءات من عينة لأخرى لذا فهى متغيرات عشوائية. ومن البديهى أن التقدير بنقطة لأى معلمة من معالم المجتمع الإحصائي لا يُتوقع أن يكون مطابقاً تماماً لقيمة المعلمة المطلوب تقديرها أي أنه من المتوقع حدوث خطأ معين فى التقدير. والفرق بين المعلمة والمقدر لها يسمى بمقدار الخطأ، والمقدر الجيد هو الذى يجعل هذا الخطأ مساوياً للصفر أو قريباً من الصفر. أما التقدير بفترة فهو تحديد فترة يتوقع أن تقع داخلها قيمة المعلمة المراد تقديرها وتقع هذه المعلمة داخل حدود الفترة بدرجة ثقة (أو باحتمال) يحدد ذلك. وتكتب الصيغة العامة للتقدير بفترة للمعلمة θ كالآتي:

$$\theta_L \le \theta \le \theta_U$$

حيث θ_L, θ_U هما حدود الفترة. ويتم تحديدهما بحيث يكون احتمال أن تقع θ بينهما مساوياً المقدار $1-\alpha$ بحيث $0 < \theta < 1$ أى أن

$$P[\theta_L \le \theta \le \theta_U] = 1 - \alpha$$

حيث تسمى الفترة التى حدها الأدنى θ_L وحدها الأعلى θ_U بأنها $100\%(1-\alpha)$، فترة ثقة للمعلمة θ، ويسمى المقدار $(1-\alpha)$ بدرجة الثقة. ومن الملاحظ أنه كلما كان طول الفترة صغيراً زادت دقة الفترة. وسوف نهتم بعرض التقدير بفترة لمعالم مجتمع إحصائي واحد.

١١-٢-١ تقدير فترة الثقة للمتوسط (Confidence Interval Estimation for the Mean)

قبل إجراء عملية التقدير يجب تحديد أكبر قيمة للخطأ فى تقدير المتوسط وأيضاً تقدير حجم العينة التى سنقوم بسحبها لإجراء عملية التقدير لمتوسط المجتمع.

تقدير أكبر قيمة للخطأ فى تقدير المتوسط:

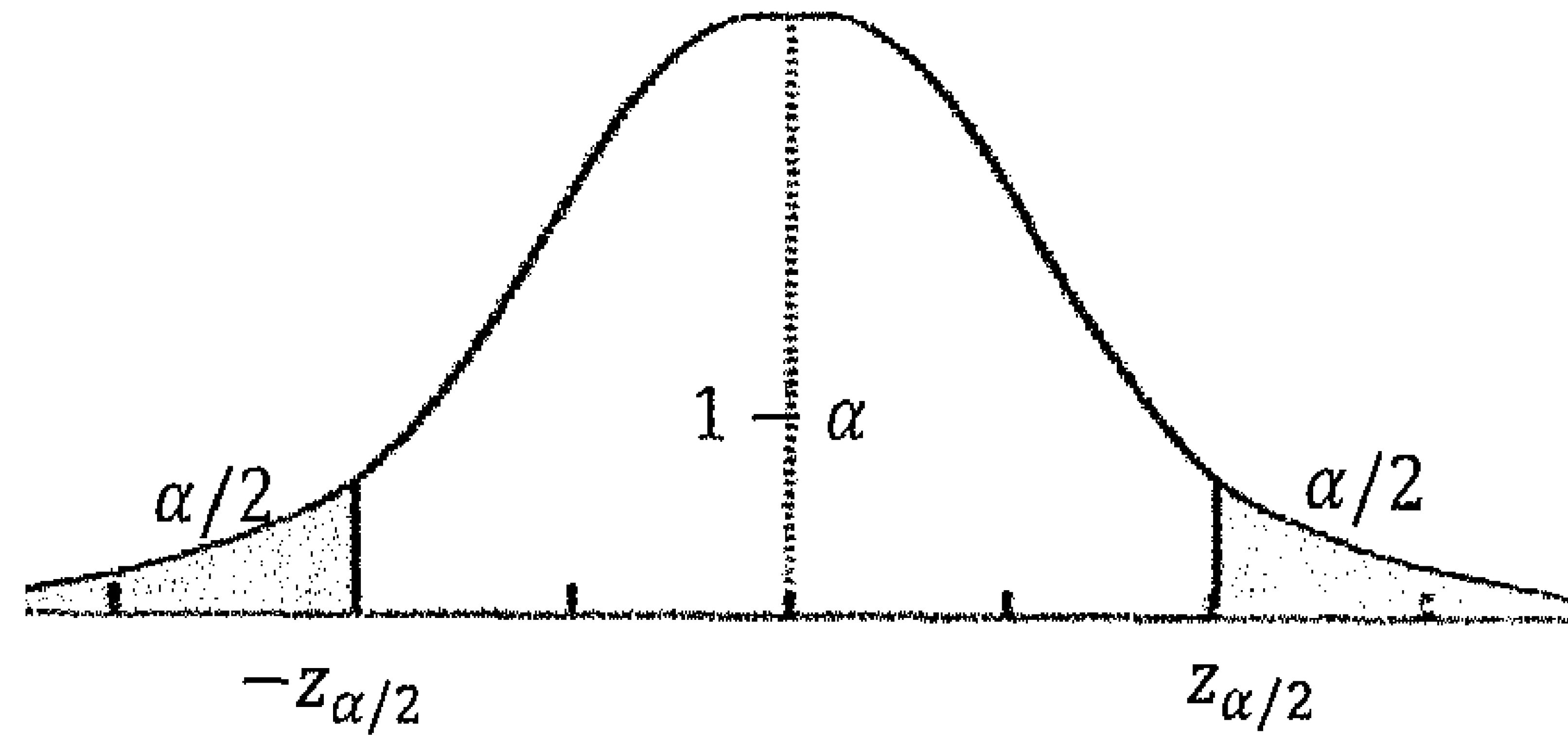
سوف تعتمد أكبر قيمة للخطأ فى تقدير المتوسط على توزيعات المعاينة لمتوسط العينة. ومن المعلوم أن الخطأ فى التقدير وليكن d هو قيمة الفرق بين المعلمة والقيمة المقدرة له أى أن:

$$d = |\mu - \hat{\mu}| = |\mu - \bar{X}|$$

ومن توزيعات المعاينة فيوجد لدينا ثلاث حالات كما يلي:

١- إذا كان المجتمع يتبع التوزيع الطبيعي $N(\mu, \sigma^2)$ وتباينه معلوم فإن المقدار $Z = \frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}$ يتبع التوزيع الطبيعي القياسي ومن ثم فإن

$$P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$



$$\Rightarrow P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} \leq z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\Rightarrow P\left(-z_{\frac{\alpha}{2}}\sigma_{\bar{x}} \leq \bar{X} - \mu \leq z_{\frac{\alpha}{2}}\sigma_{\bar{x}}\right) = 1 - \alpha$$

$$\Rightarrow P\left(|\bar{X} - \mu| \leq z_{\frac{\alpha}{2}}\sigma_{\bar{x}}\right) = 1 - \alpha$$

$$\Rightarrow P\left(|\mu - \hat{\mu}| \leq z_{\frac{\alpha}{2}}\sigma_{\bar{x}}\right) = 1 - \alpha$$

$$\Rightarrow P\left(d \leq z_{\frac{\alpha}{2}}\sigma_{\bar{x}}\right) = 1-\alpha$$

ومن ثم فإنه باحتمال $1-\alpha$ فإن الخطأ فى تقدير المتوسط لن يتعدى المقدار

$$d = z_{\frac{\alpha}{2}}\,\sigma_{\bar{x}}$$

٢- إذا كان حجم العينة كبير $(n \geq 30)$ وكان المجتمع له التوزيع الطبيعي أو غير الطبيعي فإنه من نظرية النهاية المركزية فإن المقدار $Z = \frac{\bar{X}-\mu_{\bar{x}}}{\sigma_{\bar{x}}}$ يتبع تقريباً التوزيع الطبيعي القياسي ومن ثم فإن

$$P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

وبنفس الطريقة السابقة فإنه باحتمال $1-\alpha$ فإن الخطأ فى تقدير المتوسط لن يتعدى المقدار

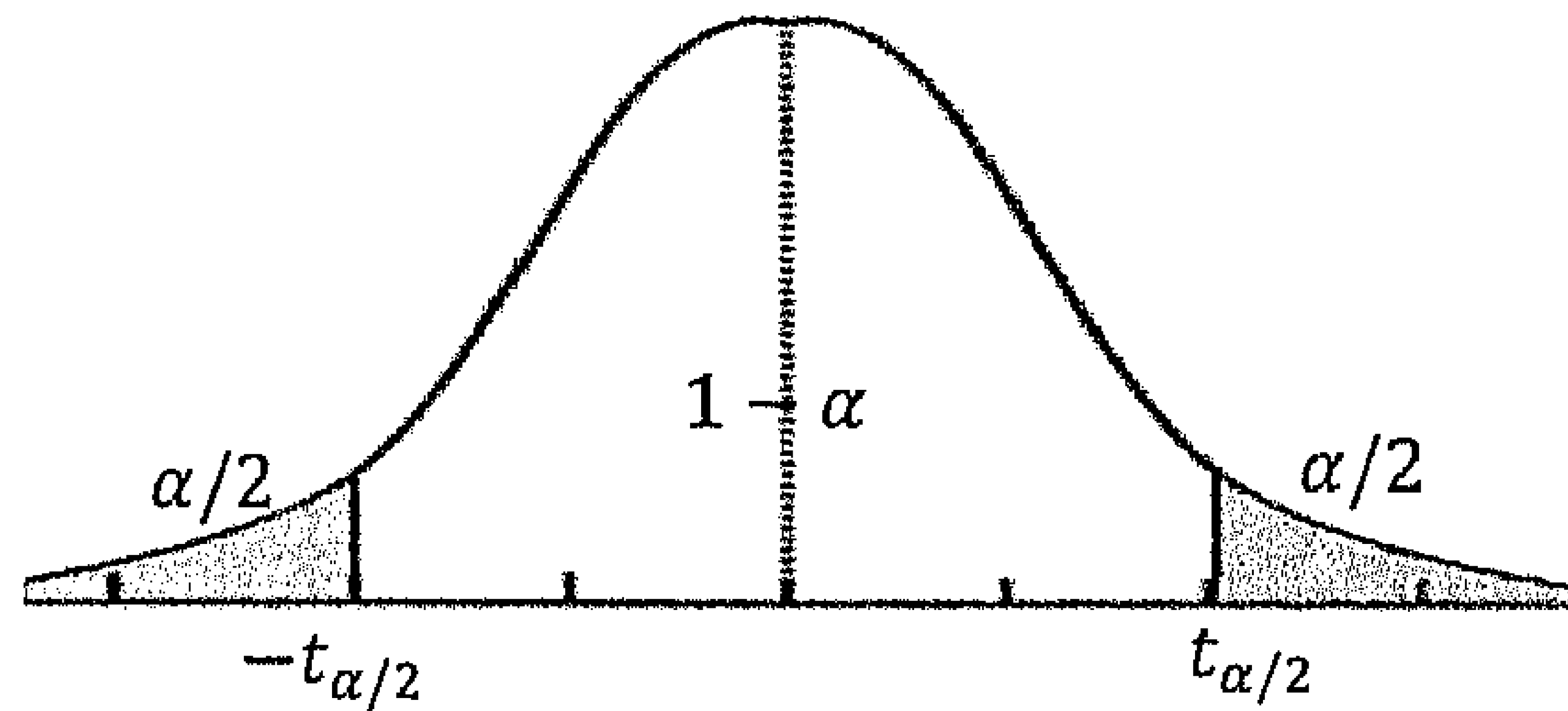
$$d = z_{\frac{\alpha}{2}}\,\sigma_{\bar{x}}$$

٣- إذا كان المجتمع يتبع التوزيع الطبيعي $N(\mu,\sigma^2)$ وتباينه غير معلوم وحجم العينة صغير $(n < 30)$ فإن

$$T = \frac{\bar{X}-\mu_{\bar{x}}}{s_{\bar{x}}} \sim t_{n-1}$$

حيث

$$P\left(-t_{\frac{\alpha}{2}} \leq T \leq t_{\frac{\alpha}{2}}\right) = 1-\alpha$$



$$\Rightarrow P\left(-t_{\frac{\alpha}{2}} \leq \frac{\bar{X}-\mu}{s_{\bar{x}}} \leq t_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Rightarrow P\left(-t_{\frac{\alpha}{2}}\,s_{\bar{x}} \leq \bar{X}-\mu \leq t_{\frac{\alpha}{2}}\,s_{\bar{x}}\right) = 1-\alpha$$

$$\Rightarrow P\left(|\bar{X}-\mu| \leq t_{\frac{\alpha}{2}}\,s_{\bar{x}}\right) = 1-\alpha$$

$$\Longrightarrow P\left(|\mu-\hat{\mu}| \leq t_{\frac{\alpha}{2}}\, s_{\bar{x}}\right)=1-\alpha$$

$$\Longrightarrow P\left(d \leq t_{\frac{\alpha}{2}}\, s_{\bar{x}}\right)=1-\alpha$$

ومن ثم فإنه باحتمال $1-\alpha$ فإن الخطأ فى تقدير المتوسط لن يتعدى المقدار

$$d=t_{\frac{\alpha}{2}}\, s_{\bar{x}}$$

ومن الملاحظ أنه فى الحالة الأولى لا نهتم بقيمة حجم العينة n كبيراً أو صغيراً. بينما فى الحالة الثانية لانهتم فيها هل التباين معلوم أم لا فإذا كان معلوماً فسوف نستخدمه وإذا كان مجهولاً سوف نستخدم تباين العينة لكن فى الحالة الثالثة نهتم فيها بالتوزيع فيجب أن يكون طبيعياً والتباين غير معلوم وأيضا حجم العينة صغيراً. حيث

$$\sigma_{\bar{x}}^{2}=\begin{cases}\dfrac{\sigma^{2}}{n}, & \text{إذا كان المجتمع الأصلى كبيراً}\\[2mm] \dfrac{\sigma^{2}}{n}\left(\dfrac{N-n}{N-1}\right), & \text{إذا كان المجتمع الأصلى محدوداً}\end{cases}$$

$$s_{\bar{x}}^{2}=\begin{cases}\dfrac{s^{2}}{n}, & \text{إذا كان المجتمع الأصلى كبيراً}\\[2mm] \dfrac{s^{2}}{n}\left(\dfrac{N-n}{N-1}\right), & \text{إذا كان المجتمع الأصلى محدوداً}\end{cases}$$

$$\Phi\left(z_{\frac{\alpha}{2}}\right)=1-\frac{\alpha}{2},$$

$$t_{\frac{\alpha}{2}}=t\left(n-1, \frac{\alpha}{2}\right).$$

مثال (١١-١)

فى دراسة عن الإنتاج بأحد المصانع كان متوسط الإنتاج اليومي خلال 60 يوماً هو 18.85 طناً. فاحسب باحتمال 0.95 مقدار القيمة العظمى للخطأ فى تقدير متوسط الإنتاج، علماً بأن الانحراف المعيارى لإنتاج المصنع هو 2.55 طناً.

الحل

من المعطيات $n = 60, \bar{x} = 18.85, 1 - \alpha = 0.95, \sigma = 2.55$ وحيث إن المجتمع غير طبيعي وحجم العينة كبير $n > 30$ فإن أكبر قيمة للخطأ هى

$$d = z_{\frac{\alpha}{2}} \sigma_{\bar{x}}$$

وحيث إن التباين معلوم والمجتمع غير محدود

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.55}{\sqrt{60}} = 0.3292$$

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$$

$$\Phi\left(z_{\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2} = 1 - \frac{0.05}{2} = 0.975,$$

ومن جدول التوزيع الطبيعي القياسي فإن

$$\Rightarrow z_{\alpha/2} = 1.96$$

$$\therefore d = z_{\frac{\alpha}{2}} \sigma_{\bar{x}} = 1.96 \times 0.3292 = 0.6452$$

مثال(١١-٢)

فى المثال السابق إذا كان عدد الأيام التى تم تشغيل المصنع فيها هو N=400 يوماً. فاحسب مقدار القيمة العظمى للخطأ فى التقدير باحتمال 0.95.

الحل

حيث إن التباين معلوم والمجتمع محدود N=400 فإن

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}} = \frac{2.55}{\sqrt{60}} \sqrt{\frac{400 - 60}{400 - 1}} = 0.30389$$

$$\therefore d = 1.96(0.30389) = 0.59562$$

تقديرحجم العينة اللازم لتقدير المتوسط:

إن تحديد حجم العينة اللازم لإجراء أى بحث أو دراسة إحصائية من الموضوعات المهمة التى ينبغى اتخاذ قرار بشأنه عند تصميم البحث، وذلك لأن اختيار عينة كبيرة يؤدى إلى زيادة في التكاليف والجهد وأيضاً الوقت كما أن اختيار عينة صغيرة قد يجعل النتائج التى تحسب منها ذات قيمة محدودة. فعند تحديد مقدار الدقة

المطلوبة (أو القيمة العظمى للخطأ) في التقدير لمتوسط المجتمع وباحتمال معين $(1-\alpha)$ فإنه يمكن تحديد حجم العينة n اللازم للحصول على ذلك الخطأ في حالتين فقط:

- إذا كان المجتمع الأصلي يتبع التوزيع الطبيعي وتباينه معلوماً وأيضاً إذا كان حجم العينة كبيراً لأن توزيع المعاينة في هاتين الحالتين سيكون التوزيع الطبيعي، أو يقترب منه والقيم الجدولية للتوزيع الطبيعي القياسى لا تعتمد على قيمة n.
- إذا كان المجتمع الأصلي يتبع التوزيع الطبيعي وتباينه غير معلوم وحجم العينة صغيراً فإن توزيع المعاينة سوف يؤول لتوزيع t بدرجة حرية $n-1$ وستكون القيمة الجدولية معتمدة على n والمطلوب أن نقوم بتقدير قيمة n المجهولة.

حجم العينة اللازم لتقدير المتوسط بدرجة ثقة $(1-\alpha)100\%$ وبخطأ لا يتعدي d يمكن تقديره كالتالي:

١- في حالة المجتمعات الكبيرة غير المحدودة (السحب بإرجاع) فإن

$$n=\left(\frac{z_{\alpha/2}\sigma}{d}\right)^2$$

٢- في حالة المجتمعات الصغيرة المحدودة (السحب بدون إرجاع) فإن

$$n=\frac{n_0}{1+\frac{n_0}{N}},$$

حيث

$$n_0=\left(\frac{z_{\alpha/2}\sigma}{d}\right)^2$$

فإذا كان تباين المجتمع غير معلوم فسوف نستخدم تباين العينة كمقدر لتباين المجتمع ويجب ملاحظة أنه إذا كان المجتمع غير طبيعى فسيكون حجم العينة التى سنحصل عليها كبيراً $(n \geq 30)$.

مثال (١١-٣)

في دراسة عن الأجور الشهرية لعمال أحد المصانع، ما عدد العمال المطلوب لتقدير متوسط الأجر الشهرى بدرجة 99% ثقة بحيث إن الخطأ في تقدير متوسط الأجر الشهرى لعمال المصنع لا يتعدى 1.6، علماً بأن الانحراف المعياري للأجور الشهرية لعمال المصنع هو 10 ريالات سعودية.

الحل

حيث إن $d = 1.6, \sigma = 10$ وأيضاً

$$1-\alpha = 0.99 \Rightarrow \alpha = 0.01$$

$$\Rightarrow \Phi\left(z_{\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2} = 1-0.005 = 0.995$$

ومن جدول التوزيع الطبيعي القياسى فإن

$$\Rightarrow z_{\alpha/2} = 2.575$$

وتقدير حجم العينة يعطي كالتالي:

$$n = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{d}\right)^2$$

$$= \left(\frac{2.575 \times 10}{1.6}\right)^2 = 259.0088 \cong 260$$

ومن الملاحظ أن المجتمع الأصلى لا يتبع التوزيع الطبيعي لذا فإن حجم العينة التى حصلنا عليها كبير.

مثال(١١-٤)

فى المثال السابق ما حجم العينة n إذا علم أن العدد الكلي لعمال المصنع هو N=1000 عاملاً.

الحل

حيث إن المجتمع محدود، $N=1000, d = 1.6, \sigma = 10$ وأيضاً $z_{\alpha/2} = 2.575$ فإن

$$n_0 = \left(\frac{z_{\frac{\alpha}{2}}\sigma}{d}\right)^2$$

$$= \left(\frac{2.575 \times 10}{1.6}\right)^2 = 259.0088$$

$$\therefore n = \frac{n_0}{1+\frac{n_0}{N}} = \frac{259.0088}{1+\frac{259.0088}{1000}} = 205.72 \cong 206$$

تقدير فترة الثقة للمتوسط:

إذا كان أكبر فرق للخطأ المطلق بين متوسط المجتمع μ ومتوسط العينة $\bar{x}$ هو المقدار الموجب d باحتمال $(1-\alpha)$ فإن:

$$P[|\mu - \bar{x}| \le d] \ge 1-\alpha$$

$$\Rightarrow P[-d \le \mu - \bar{x} \le d] \ge 1 - \alpha$$

ومنها نجد أن

$$P[\bar{x} - d \le \mu \le \bar{x} + d] \ge 1 - \alpha$$

ومن ثم فإن $(1-\alpha)100\%$ فترة ثقة للمتوسط μ هى

$$\bar{x} - d \le \mu \le \bar{x} + d$$

حيث d أكبر قيمة للخطأ فى تقدير المتوسط بدرجة ثقة $(1-\alpha)100\%$ وتختلف قيمة d كما سبق.

مثال (١١-٥)

سحبت عينة عشوائية حجمها 25 من مجتمع يتبع التوزيع الطبيعي تباينه 16، فإذا كان المتوسط الحسابي لهذه العينة هو 55 فأوجد 95% فترة ثقة لمتوسط ذلك المجتمع.

الحل

بما أن المجتمع يتبع التوزيع الطبيعي، $n = 25, \sigma^2 = 16, \bar{x} = 55, (1-\alpha)100\% = 95\%$ فإن $(1-\alpha)100\%$ فترة ثقة للمتوسط μ هى

$$\bar{x} - d \le \mu \le \bar{x} + d$$

وحيث إن المجتمع طبيعي وتباينه معلوم فإن

$$d = z_{\frac{\alpha}{2}}\, \sigma_{\bar{x}}$$

$$\Rightarrow (1-\alpha) = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$$

$$\therefore \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0.025 = 0.475$$

ومن جدول التوزيع الطبيعي القياسي نجد أن

$$z_{\frac{\alpha}{2}} = 1.96$$

وحيث إن المجتمع غير محدود فإن

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{16}{25} = 0.64 \Rightarrow \sigma_{\bar{x}} = \sqrt{0.64} = 0.8$$

$$\therefore d = z_{\frac{\alpha}{2}}\, \sigma_{\bar{x}} = 1.96 \times 0.8 = 1.568$$

وبالتالى فإن 95% فترة ثقة للمتوسط μ هى

$$55 - 1.568 \leq \mu \leq 55 + 1.568$$

$$\Longrightarrow 53.432 \leq \mu \leq 56.568$$

مثال (١١-٦)

فى المثال السابق إذا كان المجتمع يحتوى على 500 عنصراً. فأوجد 95% فترة ثقة لمتوسط ذلك المجتمع.

الحل

المعطيات هى $n = 25, \sigma^2 = 16, \bar{x} = 55, (1-\alpha)100\% = 95\%$ فإن

$$\Longrightarrow (1-\alpha) = 0.95 \Longrightarrow \alpha = 0.05 \Longrightarrow \alpha/2 = 0.025$$

$$\therefore \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0.025 = 0.475$$

ومن جدول التوزيع الطبيعي القياسي نجد أن

$$z_{\frac{\alpha}{2}} = 1.96$$

وحيث إن المجتمع محدود فإن

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{16}{25}\left(\frac{500-25}{500-1}\right) = 0.6092$$

$$\Longrightarrow \sigma_{\bar{x}} = \sqrt{0.6092} = 0.7805$$

$$\therefore d = z_{\frac{\alpha}{2}}\,\sigma_{\bar{x}} = 1.96 \times 0.7805 = 1.5298$$

ومن ثم فإن 95% فترة ثقة للمتوسط μ هى

$$\bar{x} - d \leq \mu \leq \bar{x} + d$$

$$\Longrightarrow 55 - 1.5298 \leq \mu \leq 55 + 1.5298$$

$$\Longrightarrow 53.4702 \leq \mu \leq 56.5298$$

مثال(١١-٧)

عينة حجمها 64 أعطت انحرافاً معيارياً مقداره 6 ومتوسط 60 أوجد 99% فترة ثقة لمتوسط المجتمع المسحوب منه العينة.

الحل

المعطيات هى أن المجتمع لا يتبع التوزيع الطبيعي وأن $\bar{x} = 60,\ n = 64,\ s = 6,$ فإن $(1-\alpha)100\%$ فترة ثقة للمتوسط μ هى

$$\bar{x} - d \leq \mu \leq \bar{x} + d$$

ومن الملاحظ أن المجتمع لا يتبع التوزيع الطبيعي لكن حجم العينة كبير $n = 64 > 30$ فإن

$$d = z_{\frac{\alpha}{2}}\sigma_{\bar{x}}$$

وأيضاً

$$(1-\alpha) = 0.99 \Longrightarrow \alpha = 0.01 \Longrightarrow \alpha/2 = 0.005$$

$$\Longrightarrow \Phi\left(z_{\alpha/2}\right) = 1 - \frac{\alpha}{2} = 1 - 0.005 = 0.995$$

ومن جدول التوزيع الطبيعي القياسي نجد أن

$$z_{\frac{\alpha}{2}} = 2.575$$

وحيث إن المجتمع غير محدود فإن

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{6}{\sqrt{64}} = 0.75$$

$$\Longrightarrow d = z_{\frac{\alpha}{2}}\sigma_{\bar{x}} = 2.575 \times 0.75 = 1.93125$$

ومن ثم فإن 99% فترة ثقة للمتوسط هى

$$60 - 1.93125 \leq \mu \leq 60 + 1.93125$$

$$\Longrightarrow 58.06875 \leq \mu \leq 61.93125$$

مثال(١١-٨)

إذا كان الانحراف المعياري لأطوال مجتمع الطلبة (بالسنتيمتر) هو 6.2 اخترنا عينة مكونة من 10 طلاب فكانت أطوالهم كما يلى: 179 ,176 ,174 ,173 ,170 ,169 ,168 ,165 ,163 ,163 فأوجد 95% فترة ثقة لمتوسط أطوال مجتمع الطلاب علماً بأن أطوال الطلاب يخضع للتوزيع الطبيعي.

الحل

بفرض أن X تمثل أطوال الطلاب حيث $\sigma = 6.2, n = 10, (1-\alpha)100\% = 95\%$ فإن $(1-\alpha)100\%$ فترة ثقة للمتوسط μ هى

$$\bar{x} + d \leq \mu \leq \bar{x} + d$$

وحيث إن المجتمع يتبع التوزيع الطبيعي وتباينه معلوم فإن

$$d = z_{\frac{\alpha}{2}}\sigma_{\bar{x}}$$

وأيضاً

$$(1-\alpha) = 0.95 \Longrightarrow \alpha = 0.05 \Longrightarrow \alpha/2 = 0.025$$

$$\Longrightarrow \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975$$

ومن جدول التوزيع الطبيعي القياسي فإن

$$z_{\frac{\alpha}{2}} = 1.96$$

وحيث إن المجتمع غير محدود فإن

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6.2}{\sqrt{10}} = 1.96061$$

$$\Longrightarrow d = z_{\frac{\alpha}{2}}\sigma_{\bar{x}} = 1.96 \times 1.96061 = 3.84279$$

والوسط الحسابي للعينة يمكن تعيينه كما يلي:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}[\,163 + 163 + \cdots + 179]$$

$$= \frac{1700}{10} = 170$$

ومن ثم فإن 95% فترة ثقة للمتوسط هي

$$170 - 3.8429 \le \mu \le 170 + 3.84279$$

$$\Longrightarrow 166.1571 \le \mu \le 173.84279$$

مثال(١١-٩)

أخذت عينة عشوائية مكونة من 9 أفدنة مزروعة بنوع معين من القمح فوجد أن إنتاجها بالإردب كالاتى: 12 ,10 ,11 ,9 ,8 ,10.5 ,9.5 ,11.5 ,8.5 فاحسب 95% فترة ثقة لمتوسط إنتاجية الفدان من ذلك النوع من القمح وذلك بفرض أن إنتاجية الفدان من ذلك النوع من القمح لها التوزيع الطبيعي.

الحل

بفرض أن X تمثل إنتاجية الفدان من القمح، $n = 9, (1-\alpha)100\% = 95\%$ فإن $(1-\alpha)100\%$ فترة ثقة للمتوسط μ هي

$$\bar{x} - d \le \mu \le \bar{x} + d$$

وحيث إن المجتمع يتبع التوزيع الطبيعي وتباينه غير معلوم

$$d = t_{\frac{\alpha}{2}} s_{\bar{x}}$$

$$(1-\alpha) = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$$

ومن جدول توزيع t فإن

$$\Rightarrow t_{\alpha/2} = t\left(n-1, \frac{\alpha}{2}\right) = t(8, 0.025) = 2.306$$

ولتعيين الوسط الحسابي والتباين للعينة نقوم بتكوين الجدول التالى:

x_i	x_i^2
8.5	72.25
11.5	132.25
9.5	90.25
10.5	110.25
8.0	64.00
9.0	81.00
11.0	121.00
10.0	100.00
12.0	144.00
$\sum x = 90$	$\sum x^2 = 915$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{90}{9} = 10$$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right]$$

$$= \frac{1}{9-1}[915 - 9(10)^2] = 1.875$$

$$\Rightarrow s = \sqrt{1.875} = 1.3693$$

وحيث إن المجتمع غير محدود فإن

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1.3693}{\sqrt{9}} = 0.4564$$

$$\Rightarrow d = t_{\frac{\alpha}{2}} s_{\bar{x}} = 2.306 \times 0.4564 = 1.05254$$

95% فترة ثقة للمتوسط هى

$$10 - 1.05254 \leq \mu \leq 10 + 1.05254$$

$$\Rightarrow 8.94746 \leq \mu \leq 11.05254$$

مثال(١١-١٠)

أخذت عينة عشوائية مكونة من 500 فدان من منطقة ما بها 60000 فداناً مزروعة ذرة، فوجد أن المتوسط الحسابى والانحراف المعيارى المحسوبين من العينة هما 10 ,2 إردب على الترتيب. فاحسب

(أ) 95% فترة ثقة لمتوسط إنتاجية الفدان بتلك المنطقة.

(ب) 95% فترة ثقة لإجمالي إنتاج الذرة بتلك المنطقة.

الحل

المعطيات $n = 500, N = 60000, \bar{x} = 10, \; s = 2, (1-\alpha)100\% = 95\%$

(أ) بالنسبة لمتوسط إنتاجية الفدان:

$(1-\alpha)100\%$ فترة ثقة للمتوسط μ هى

$$\bar{x} - d \leq \mu \leq \bar{x} + d$$

وحيث إن المجتمع لا يتبع التوزيع الطبيعي، لكن حجم العينة كبير $n = 500 > 30$ ولذلك فإن

$$d = z_{\frac{\alpha}{2}} \, s_{\bar{x}}$$

وأيضاً

$$(1-\alpha) = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$$

$$\Rightarrow \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975$$

ومن جدول التوزيع الطبيعي القياسى فنجد أن

$$z_{\frac{\alpha}{2}} = 1.96$$

وحيث إن المجتمع محدود فإن

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{2}{\sqrt{500}}\sqrt{\frac{60000-500}{60000-1}} = 0.08907$$

$$\Rightarrow d = z_{\frac{\alpha}{2}} s_{\bar{x}} = 1.96 \times 0.08907 = 0.17458$$

ومن ثم فإن 95% فترة ثقة للمتوسط هى

$$10 - 0.17458 \leq \mu \leq 10 + 0.17458$$

$$\Rightarrow 9.82542 \leq \mu \leq 10.17458$$

(ب) بالنسبة لإجمالي انتاجية الفدان:

حيث إن إجمالي إنتاجية الفدان فى تلك المنطقة هو $\sum_{i=1}^{N} X_i$ لكن من قانون المتوسط فإن:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \Rightarrow \sum_{i=1}^{N} X_i = N\mu$$

فإن إجمالي إنتاجية الفدان فى تلك المنطقة هو $N\mu$ ومن ثم فإن فترة الثقة لإجمالي إنتاجية الفدان هى حاصل ضرب فترة الثقة لمتوسط انتاجية الفدان فى تلك المنطقة فى عدد الأفدنه(حجم المجتمع) كالتالي:

$$\Rightarrow 60000 \times 9.82542 \leq N\,\mu \leq 60000 \times 10.17458$$

$$\Rightarrow 589525.2 \leq N\,\mu \leq 610474.8$$

١١-٢-٢ تقدير فترة الثقة للنسبة (Confidence Interval Estimation for the Proportion)

تواجه الباحثين فى كثير من المشاكل الطبيعية مشكلة تقدير بعض النسب مثل نسبة النجاح فى أحد الاختبارات، نسبة المصابين بمرض معين، نسبة الوفيات بسبب مرض ما.

تقدير أكبر قيمة للخطأ d فى تقدير نسبة ظاهرة ما:

إذا كانت نسبة ظاهرة ما فى المجتمع هى p وقمنا باختيار عينة حجمها n (حيث $n \geq 30$) من ذلك المجتمع، فإذا كان عدد الأفراد فى العينة الذين لديهم خاصية معينة هى x ولنفرض أن النسبة لنفس الظاهرة فى العينة هى r حيث $r = \frac{x}{n}$ فإنه من توزيع المعاينة للنسبة r فإن المتغير $\frac{r-\mu_r}{\sigma_r}$ تتبع التوزيع الطبيعي القياسي عندما يكون حجم العينة كبيراً، وكما سبق مع المتوسط فإن أكبر قيمة للخطأ فى تقدير النسبة p باحتمال $(1-\alpha)$ هى

$$d = z_{\frac{\alpha}{2}}\,\sigma_{\hat{p}}$$

حيث

$$\sigma_{\hat{p}}^2 = \begin{cases} \dfrac{\hat{p}(1-\hat{p})}{n}, & \text{اذا كان المجتمع غير محدود} \\ \dfrac{\hat{p}(1-\hat{p})}{n}\left(\dfrac{N-n}{N-1}\right), & \text{اذا كان المجتمع محدوداً} \end{cases}$$

ومن المعلوم أن $\hat{p}$ هي المقدر بنقطة لنسبة الظاهرة فى المجتمع وهي تمثل النسبة فى العينة(لنفس الظاهرة) المختارة من المجتمع فإن

$$\hat{p} = r = \frac{x}{n}$$

تقدير حجم العينة اللازم لتقدير النسبة:

حجم العينة اللازم لتقدير النسبة p بخطأ لا يتعدى d وباحتمال $(1-\alpha)$ يعطى كما يلي:

١- فى حالة المجتمعات غير المحدودة هى

$$n = \left(\frac{z_{\alpha/2}}{d}\right)^2 \hat{p}(1-\hat{p})$$

٢- فى حالة المجتمعات المحدودة فإن

$$n = \frac{k_0}{1 + \dfrac{k_0 - 1}{n}}$$

حيث

$$k_0 = \left(\frac{z_{\alpha/2}}{d}\right)^2 \hat{p}(1-\hat{p}),$$

وحيث إن $\hat{p}$ لا تكون معلومة قبل سحب العينة فإنها تستبدل بأى قيمة تقريبية لها نحصل عليها من عينة مسبقة تستعمل لهذا الغرض أو من معلومات سابقة تعطى قيمة تقريبية لها أو بمعلومية p تقع فى فترة معينة وفى هذه الحالة نختار أقرب قيمة إلى 0.5 وإذا لم يكن هناك أى معلومات عن $\hat{p}$ فإننا نعوض عنها بالقيمة 0.5 مع ملاحظة أن $0 \le p \le 1$.

مثال(١١-١١)

احسب حجم العينة العشوائية البسيطة اللازم سحبها من مجتمع معين لتقدير نسبة المدخنين فى ذلك المجتمع بخطأ لا يتعدى 0.019 وبدرجة ثقة 95% إذا علم أن تلك النسبة تقع بين 45%, 60%.

الحل

من البيانات المعطاة فإن $d = 0.19, 1-\alpha = 0.95$. حيث إن المجتمع غير محدود فإن

$$n = \left(\frac{z_{\alpha/2}}{d}\right)^2 \hat{p}(1-\hat{p})$$

وواضح أن أقرب قيمة فى الفترة المعطاة 0.5 هى $\hat{p} = 0.45$ وأن

$$1-\alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$$
$$\Rightarrow \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975$$

ومن جدول التوزيع الطبيعي القياسي فإن

$$z_{\alpha/2} = 1.96$$

$$\therefore n = \left(\frac{1.96}{0.019}\right)^2 0.45(1-0.45) = 2633.78 \cong 2634$$

مثال(١١-١٢)

احسب حجم العينة العشوائية البسيطه اللازم سحبها من مجتمع لتقدير نسبة الأفراد الموظفين فى ذلك المجتمع. بخطأ لايتعدى 0.04 وبدرجة ثقة 99% وبفرض أن تلك النسبة تقع بين 70%, 85%.

الحل

من البيانات المعطاة فإن $d = 0.04, (1-\alpha) = 0.99$. حيث إن المجتمع غير محدود فإن

$$n = \left(\frac{z_{\alpha/2}}{d}\right)^2 \hat{p}(1-\hat{p})$$

وواضح أن أقرب قيمة فى الفترة المعطاة من 0.5 هى $\hat{p} = 0.7$ وأن

$$1-\alpha = 0.99 \Rightarrow \alpha = 0.01 \Rightarrow \frac{\alpha}{2} = 0.005$$
$$\Rightarrow \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0.005 = 0.995$$

ومن جدول التوزيع الطبيعي القياسي فإن

$$z_{\alpha/2} = 2.575$$

$$\therefore n = \left(\frac{2.575}{0.04}\right)^2 \times 0.7(1-0.7) = 870.27 \cong 871$$

مثال(١١-١٣)

من مجتمع مكون من 10000 شخصاً احسب حجم العينة اللازمة لتقدير نسبة المدخنين فى ذلك المجتمع بخطأ لا يتعدى 0.035 وبدرجة ثقة 95%.

الحل

من البيانات المعطاة فإن $N = 10000, d = 0.035, (1-\alpha) = 0.95$. حيث إن المجتمع محدود فإن حجم العينة هو

$$n = \frac{k_0}{1 + \frac{k_0 - 1}{N}},$$

حيث

$$k_0 = \left(\frac{z_{\alpha/2}}{d}\right)^2 \hat{p}(1-\hat{p})$$

وحيث أنه لا توجد لدينا أي معلومات عن p لذا سوف نختار $\hat{p} = 0.5$ وأن

$$1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$$
$$\Rightarrow \Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975$$

ومن جدول التوزيع الطبيعي القياسي فإن

$$z_{\alpha/2} = 1.96$$

$$\Rightarrow k_0 = \left(\frac{1.96}{0.035}\right)^2 \times 0.5(1-0.5) = 784$$

$$\therefore n = \frac{784}{1 + \frac{784-1}{10000}} = 727.07 \cong 728$$

تقدير فترة الثقة للنسبة:

يمكن تقدير فترة الثقة لنسبة ظاهرة (صفة) ما فى المجتمع ولكن كما سبق فى توزيع المعاينة يجب أن تكون العينة التى سيتم سحبها لاستخدامها فى تقدير النسبة فى المجتمع كبيرة وذلك فإن $(1-\alpha)100\%$ فتر ثقة للنسبة فى المجتمع p هى

$$\hat{p} - d \le p \le \hat{p} + d$$

مثال(١١-١٤)

أجري بحث لتقدير نسبة الذكور فى مدينة ما فأخذت عينة من 400 فرداً، فوجد أن منهم 164 من الذكور فاحسب 95% فترة ثقة لنسبة الذكور فى هذه المدينة.

الحل

بفرض أن x تمثل عدد الذكور فى العينة، $x = 164, n = 400, (1-\alpha)100\% = 95\%$ فإن $(1-\alpha)100\%$ فترة ثقة للنسبة p هى

$$\hat{p} - d \leq p \leq \hat{p} + d$$

حيث

$$\hat{p} = \frac{x}{n} = \frac{164}{400} = 0.41,$$

وأيضاً

$$d = z_{\alpha/2}\sigma_{\hat{p}}$$

$$(1-\alpha)100\% = 95\% \Rightarrow 1-\alpha = 0.95 \ \Rightarrow \alpha = 0.05$$

$$\Rightarrow \Phi(z_{\alpha/2}) = 1 - \alpha/2 = 1 - 0.025 = 0.975$$

ومن جدول التوزيع الطبيعي القياسي فإن

$$z_{\alpha/2} = 1.96$$

وحيث إن المجتمع غير محدود فإن

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.41(1-0.41)}{400}} = 0.0246$$

$$\Rightarrow d = z_{\alpha/2}\sigma_{\hat{p}} = 1.96 \times 0.0246 = 0.0482$$

ومن ثم فإن 95% فترة ثقة لنسبة الذكور هى

$$0.41 - 0.0482 \leq p \leq 0.41 + 0.0482$$

$$\Rightarrow 0.3618 \leq p \leq 0.4582$$

مثال(١١-١٥)

سحبت عينة عشوائية حجمها 800 مصباحاً من 30000 مصباحاً والذى يمثل الإنتاج اليومي لمصنع لإنتاج المصابيح الكهربائية فوجد أن عدد المصابيح التالفة فى العينة هو 240 مصباحاً احسب:

(أ) 95% فترة ثقة لنسبة المصابيح التالفة للإنتاج اليومي للمصنع.

(ب) 95% فترة ثقة لإجمالي عدد المصابيح التالفة التى ينتجها المصنع فى اليوم الواحد.

الحل

بفرض أن x تمثل عدد المصابيح التالفة فى العينة،

$$x = 240, n = 800, N = 30000, (1-\alpha)100\% = 95\%$$

(أ) 95% فترة ثقة لنسبة المصابيح التالفة للإنتاج اليومي للمصنع

$(1-\alpha)100\%$ فترة ثقة للنسبة p هى

$$\hat{p} - d \le p \le \hat{p} + d$$

حيث

$$\hat{p} = \frac{x}{n} = \frac{240}{800} = 0.3,$$

وأيضاً

$$d = z_{\alpha/2}\sigma_{\hat{p}},$$

فإن

$$(1-\alpha)100\% = 95\% \Longrightarrow 1-\alpha = 0.95 \Longrightarrow \alpha = 0.05$$

$$\Longrightarrow \Phi(z_{\alpha/2}) = 1 - \alpha/2 = 1 - 0.025 = 0.975$$

ومن جدول التوزيع الطبيعي القياسي فإن

$$z_{\alpha/2} = 1.96$$

وحيث إن المجتمع محدود فإن

$$\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}\left(\frac{N-n}{N-1}\right)}$$

$$= \sqrt{\frac{0.3(1-0.3)}{800}\left(\frac{30000-800}{30000-1}\right)} = 0.01598$$

$$\therefore d = z_{\alpha/2}\sigma_{\hat{p}} = 1.96 \times 0.01598 = 0.03133$$

ومن ثم فإن 95% فترة ثقة لنسبة المصابيح التالفة فى الإنتاج اليومي للمصنع هى

$$0.3 - 0.0313 \le p \le 0.3 + 0.0313$$

$$\Longrightarrow 0.2687 \le p \le 0.3313$$

(ب) 95% فترة ثقة لإجمالي عدد المصابيح التالفة التى ينتجها المصنع فى اليوم الواحد

من المعلوم أن نسبة المصابيح التالفة في المجتمع هى ناتج قسمة عدد المصابيح التالفة على حجم(عدد عناصر) المجتمع بمعني

$$p = \frac{X}{N} \Longrightarrow X = Np$$

فإن فترة الثقة لإجمالي عدد المصابيح التالفه سوف نحصل عليها بضرب فترة الثقة للنسبة فى حجم المجتمع كما يلي:

$$30000 \times 0.2687 \le Np \le 30000 \times 0.3313$$

$$\Longrightarrow 8061 \le Np \le 9939$$

١١-٢-٣ تقدير فترة الثقة للتباين (Confidence Interval Estimation for the Variance)

يعتبر التباين من مقاييس التشتت التى تستخدم لدراسة مدى تجانس المجتمع، لذا فإنه من المقاييس المهمة التى نحتاج إليها ولكن فى بعض الأحيان يكون تباين المجتمع مجهولاً لذا يجب تقديره ويمكننا إجراء ذلك إذا كان لدينا مجتمعٌ طبيعيٌ. فإذا كان لدينا عينة عشوائية حجمها n مسحوبة من مجتمع يتبع التوزيع الطبيعي وتباينه σ^2 مجهول فإن $100(1-\alpha)\%$ فترة ثقة لتباين المجتمع σ^2 هى:

$$\frac{(n-1)s^2}{\chi^2(\alpha/2, n-1)} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2(1-\alpha/2, n-1)}$$

حيث s^2 هو التباين المحسوب من العينة، $\chi^2(\alpha, \nu)$ هى القيمة الجدولية لتوزيع مربع كاي عند درجة حرية $\nu = n-1$

مثال(١١-١٦)

أخذت عينة عشوائية حجمها 10 من مجتمع يتبع التوزيع الطبيعي فوجد أن تباينها هو 117.12 فأوجد 95% فترة ثقة لتباين هذا المجتمع.

الحل

من البيانات فإن $n = 10, s^2 = 117.12, (1-\alpha)100\% = 95\%$، $(1-\alpha)100\%$ فترة ثقة لتباين المجتمع σ^2 هى

$$\frac{(n-1)s^2}{\chi^2(\alpha/2, n-1)} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2(1-\alpha/2, n-1)}$$

حيث

$$(1-\alpha)100\% = 95\% \Rightarrow 1-\alpha = 0.95 \Rightarrow \alpha = 0.05$$

$$\Rightarrow \chi^2\left(\frac{\alpha}{2}, n-1\right) = \chi^2(0.025,9) = 19.02$$

$$\Rightarrow \chi^2\left(1-\frac{\alpha}{2}, n-1\right) = \chi^2(0.975,9) = 2.70$$

ومن ثم فإن 95% فترة ثقة لتباين المجتمع σ^2 هى

$$\frac{(10-1)\times 117.12}{19.02} \le \sigma^2 \le \frac{(10-1)\times 117.12}{2.7}$$

$$\Rightarrow 55.41956 \le \sigma^2 \le 390.40$$

١١-٣ اختبارات الفروض الإحصائية (Statistical Testing of Hypotheses)

يتعرض الإنسان في كثير من الحالات وفى مجالات العمل المختلفة إلى مواقف معينة تتطلب منه اتخاذ قرار بناء على معلومات محسوبة من عينة. وطبيعي أن يتخذ القرار بشيء من الحكمة وبأقل قدر ممكن من المخاطر إذ إن هذا القرار قد يترتب عليه نفقات قد تكون طائلة، وبالتالي لابد أن يكون لها ما يبررها. ولإمكانية معالجة هذا إحصائياً يجب تنفيذ الخطوات التالية بالترتيب كما يلي:

١- صياغة الفروض الإحصائية.

٢- تحديد مستوى المعنوية.

٣- تعيين إحصاء الاختبار.

٤- تحديد المناطق الحرجة.

٥- اتخاذ القرار .

وسوف نتناول كل خطوة من هذه الخطوات بشيء من التفصيل.

صياغة الفرض الإحصائي:

يمكن تعريف الفرض الإحصائي على أنه ادعاء أو تخمين حول أحد معالم المجتمع الإحصائي. ويكون المطلوب من الباحث اختبار صحة هذا الادعاء أو التخمين. ويوجد نوعين من الفروض الإحصائية.

النوع الأول من الفروض يسمي بفرض العدم (null hypothesis) وعادة يصاغ بأنه لا يوجد فرق أو عدم وجود علاقة أو عدم وجود تغيير ويرمز له بالرمز H_0.

والنوع الثاني من الفروض الإحصائية هو الفرض البديل (alternative hypothesis) وهو الفرض الذي يجب أن يكون صحيحاً إذا كان فرض العدم غير صحيح ويرمز له بالرمز H_1.

سوف نجري هنا الاختبار الإحصائي وتكون نتيجة الاختبار هو قبول فرض العدم H_0 أو رفض فرض العدم H_0 فإذا كان القرار بقبول فرض العدم فإن هذا يعنى أنه لا يوجد اختلاف أو لا يوجد فرق وأن الاختلاف الموجود لدينا هو اختلاف ظاهري نتيجة الصدفة وحدها. وقبول فرض العدم H_0 لا يعنى أنه صحيح ولكن ذلك يعنى أنه لا يوجد شواهد لدينا تعطينا غير ذلك ورفض فرض العدم يعنى أنه فرض غير صحيح. قبول أو رفض فرض العدم يعتمد على دراسة عينة عشوائية فنستخدم المعلومات التي نحصل عليها من تلك العينة لنقرر ما إذا كان فرض العدم صحيحاً أم خطأ؟ وبيانات العينة التي لا تتفق مع فرض العدم تقودنا إلى رفضه بينما البيانات التي تؤيد فرض العدم تقودنا إلى قبوله. وبصفة عامة إذا كان المطلوب اختبار فرض حول أحد معالم المجتمع الإحصائي θ أو دالة فيها فيمكن صياغة فرض العدم على الصورة التالية:

$$H_0: \theta = \theta_0$$

حيث θ_0 هي قيمة محددة للمعلم θ ويمكن صياغة الفرض البديل في إحدي الصور التالية:

$$H_1: \begin{cases} \theta \neq \theta_0 \\ \theta > \theta_0 \\ \theta < \theta_0 \end{cases}$$

مستوى المعنوية:

إذا كنا نريد اختبار فرض العدم H_0 ضد الفرض البديل H_1 فنجد أننا أمام واحدة من الحالات التالية:

(أ) أن يكون فرض العدم صحيحاً ويكون القرار بقبوله ونجد أن هذا قرار صحيح.

(ب) أن يكون فرض العدم صحيحاً ويكون القرار برفضه وهذا قرار غير صحيح وبذلك نكون قد وقعنا في خطأ ويسمى هذا النوع بأنه خطأ من النوع الأول (type I error).

(ج) أن يكون فرض العدم خطأ ويكون القرار برفضه وهذا قرار صحيح.

(د) أن يكون فرض العدم خطأ ويكون القرار بقبوله وهذا قرار غير صحيح وبذلك نكون قد وقعنا في خطأ ويسمى هذا النوع بأنه خطأ من النوع الثاني (type II error).

ويمكن تلخيص هذه الحالات في الجدول التالي:

فرض العدم / القرار	H_0 صحيح	H_0 خطأ
قبول H_0	قرار صحيح	خطأ من النوع الثاني
رفض H_0	خطأ من النوع الأول	قرار صحيح

ونجد أننا قد وقعنا في نوعين من الخطأ الأول هو رفض فرض العدم وهو صحيح ويسمى خطأ من النوع الأول، والثاني قبول فرض العدم وهو غير صحيح ويسمي خطأ من النوع الثاني ونجد أنه يفضل الوقوع في الخطأ من النوع الأول مما يؤدى إلى ثقة في القرارات التي يتم فيها قبول فرض العدم حيث يسمى احتمال الوقوع في خطأ من النوع الأول بمستوى المعنوية (level of significance) والتي يرمز لها بالرمز α والتي تسمى أحيانا بحجم منطقة الرفض (size of rejection region). لكن احتمال الوقوع في خطأ من النوع الثاني يرمز له بالرمز β

$$\alpha = P[\text{الوقوع فى خطأ من النوع الأول}]$$

$$\beta = P[\text{الوقوع فى خطأ من النوع الثانى}]$$

إن الاختبار الإحصائي الجيد هو الذي يجعل قيمة كل من α, β أصغر ما يمكن، ومن الصعب الحصول على هذا الاختبار لأنه لا يمكن تصغير كل من β, α في آن واحد حيث إن تصغير إحداهما يؤدى إلى تكبير الأخرى. لذلك فقد لجأ الإحصائيون إلى تحديد مستوى المعنوية α عند قيمة معينة ثم اختيار الاختبار الإحصائي الذي يجعل β أصغر ما يمكن. ومن قيم α الشائعة الاستعمال 0.01, 0.05 فعلى سبيل المثال إذا استخدمنا مستوى المعنوية 0.05 فهذا يعنى أن احتمال الوقوع في خطأ من النوع الأول (احتمال رفض فرض العدم وهو صحيح) هو 0.05، وهذا يعنى أيضاً أنه في المتوسط من بين كل 100 حالة (قرار) يكون منها 5 حالات فيها القرار غير صحيح وفي 95 حالة يكون القرار فيها صحيح. لذلك فإن المقدار $(1-\alpha)$ يسمى بدرجة الثقة.

تعيين إحصاء الاختبار:

لإجراء الاختبار الإحصائي وتعيين إحصاء الاختبار فإننا نقوم باتباع الخطوات التالية:

(أ) بفرض أن لدينا مجتمعاً ما يتبع توزيعاً احتمالياً معيناً وأن هذا التوزيع يعتمد على بعض المعالم.

(ب) نفرض أن المطلوب اختبار فرض حول أحد معالم هذا المجتمع أو حول دالة في هذه المعالم وليكن فرض العدم والفرض البديل على الصورة التالية:

$$H_0: \theta = \theta_0$$

$$H_1: \begin{cases} \theta \neq \theta_0 \\ \theta < \theta_0 \\ \theta > \theta_0 \end{cases}$$

(ج) فإننا سوف نقوم باختيار عينة عشوائية من ذلك المجتمع ونحاول دراسة نفس الظاهرة في العينة ونحاول تعيين الإحصاء المناظر في العينة لمعلمة المجتمع θ المراد اختبار فرض حولها.

(د) بناء على توزيعات المعاينة للإحصاء نحدد إحصاء وليكن W ويسمى إحصاء الاختبار. ومن الملاحظ أنه سيحتوى على المقدر للمعلمة θ التي يدور حولها فرض العدم مع اعتبار أن فرض العدم صحيح.

تحديد المناطق الحرجة:

المناطق الحرجة هي منطقة القبول (acceptance region) ومناطق الرفض (rejection regions) حيث إن منطقة الرفض هي المنطقة التي إذا وقعت بها قيمة إحصاء الاختبار سوف نرفض فرض العدم ومنطقة القبول هي المنطقة التي إذا وقعت بها قيمة إحصاء الاختبار سوف نقبل فرض العدم. فبناء على مستوى المعنوية α وعلى الفرض البديل يمكن تقسيم محور الإحصاء إلى منطقتين إحداهما تسمى منطقة القبول والأخرى تسمى منطقة الرفض على منحنى التوزيع الاحتمالي لإحصاء الاختبار. فإذا كان إحصاء الاختبار يتبع التوزيع الطبيعي القياسي فإننا سوف نستخدم منحنى التوزيع الطبيعي القياسي وإذا كان يتبع توزيع t سوف نستخدم منحنى توزيع t لتعيين المناطق الحرجة.

القرار:

بعد تعيين قيمة إحصاء الاختبار وتحديد مناطق الرفض والقبول فإننا سوف نأخذ قرار برفض أو قبول فرض العدم، فإذا كانت قيمة إحصاء الاختبار تقع في منطقة القبول سوف نقبل فرض العدم وإذا كانت تقع في منطقة الرفض سوف نرفض فرض العدم ونقبل الفرض البديل.

١١-٣-١ اختبار فرض حول متوسط المجتمع

يمكن تطبيق الخطوات السابقة لاختبار أي فرض إحصائي ففي حالة مجتمع واحد سوف نختبر الادعاء حول متوسط هذا المجتمع بأن قيمة متوسط المجتمع تساوى قيمة معينة أو تختلف عنها أو أقل أو أكبر من تلك القيمة وسوف نفرض إن تلك القيمة هي μ_0 وسوف نقوم بإجراء هذا الاختبار باتباع الخطوات التالية:

١- الفروض الإحصائية

فرض العدم:

$$H_0: \mu = \mu_0$$

الفرض البديل:

$$H_1: \begin{cases} \mu \neq \mu_0 \\ \mu > \mu_0 \\ \mu < \mu_0 \end{cases}$$

٢- مستوى المعنوية α

٣- إحصاء الاختبار:

- إذا كان المجتمع يتبع التوزيع الطبيعي وتباينه σ^2 معلوم فإن إحصاء الاختبار هو:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \sim N(0,1)$$

- إذا كان حجم العينة كبيراً $n \geq 30$ فإن إحصاء الاختبار هو:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} \sim N(0,1)$$

- إذا كان المجتمع يتبع التوزيع الطبيعي وتباينه σ^2 غير معلوم وحجم العينة صغيراً $(n < 30)$ فإن إحصاء الاختبار هو:

$$T = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} \sim t_{(n-1)}$$

حيث

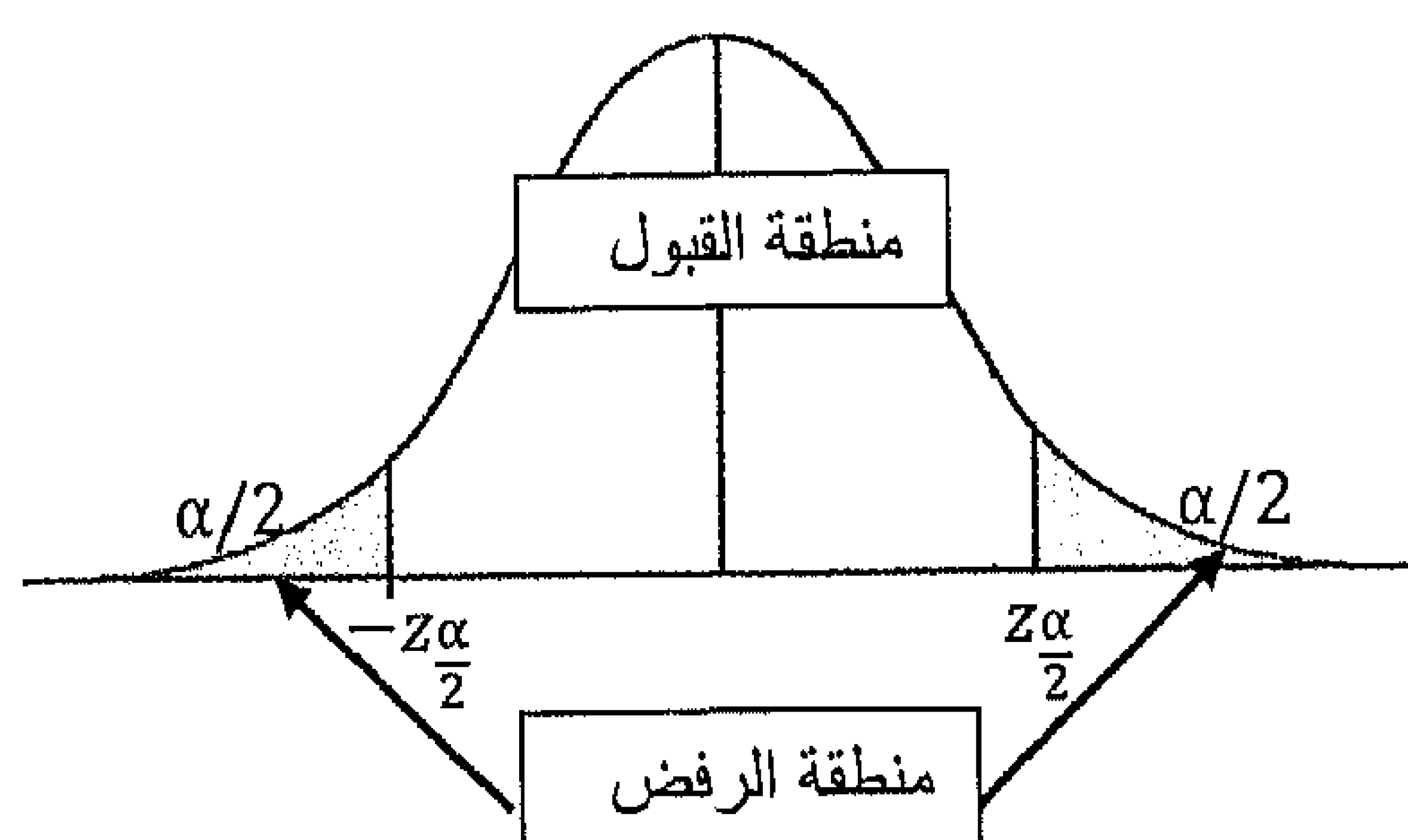
$$\sigma_{\bar{x}}^2 = \begin{cases} \dfrac{\sigma^2}{n}, & \text{اذا كان المجتمع غير محدود} \\ \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right), & \text{اذا كان المجتمع محدود} \end{cases}$$

$$s_{\bar{x}}^2 = \begin{cases} \dfrac{s^2}{n}, & \text{اذا كان المجتمع غير محدود} \\ \dfrac{s^2}{n}\left(\dfrac{N-n}{N-1}\right), & \text{اذا كان المجتمع محدود} \end{cases}$$

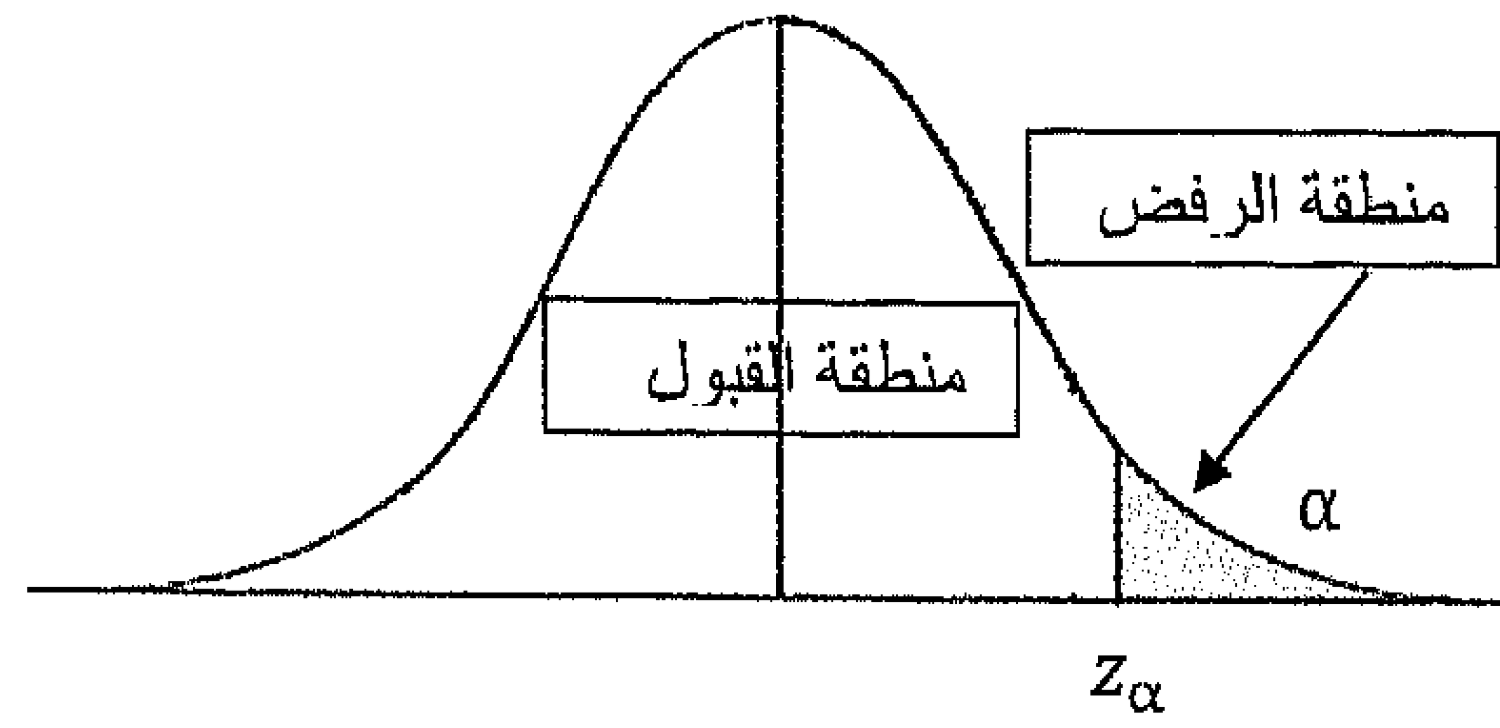
٤- تحديد المناطق الحرجة:

الحالة الأولى: إذا كان إحصاء الاختبار هو Z فإننا سوف نستخدم منحنى التوزيع الطبيعي القياسي لتحديد المناطق الحرجة، سوف نستخدم جدول التوزيع الطبيعي القياسي لتعيين حدود المناطق الحرجة بحيث:

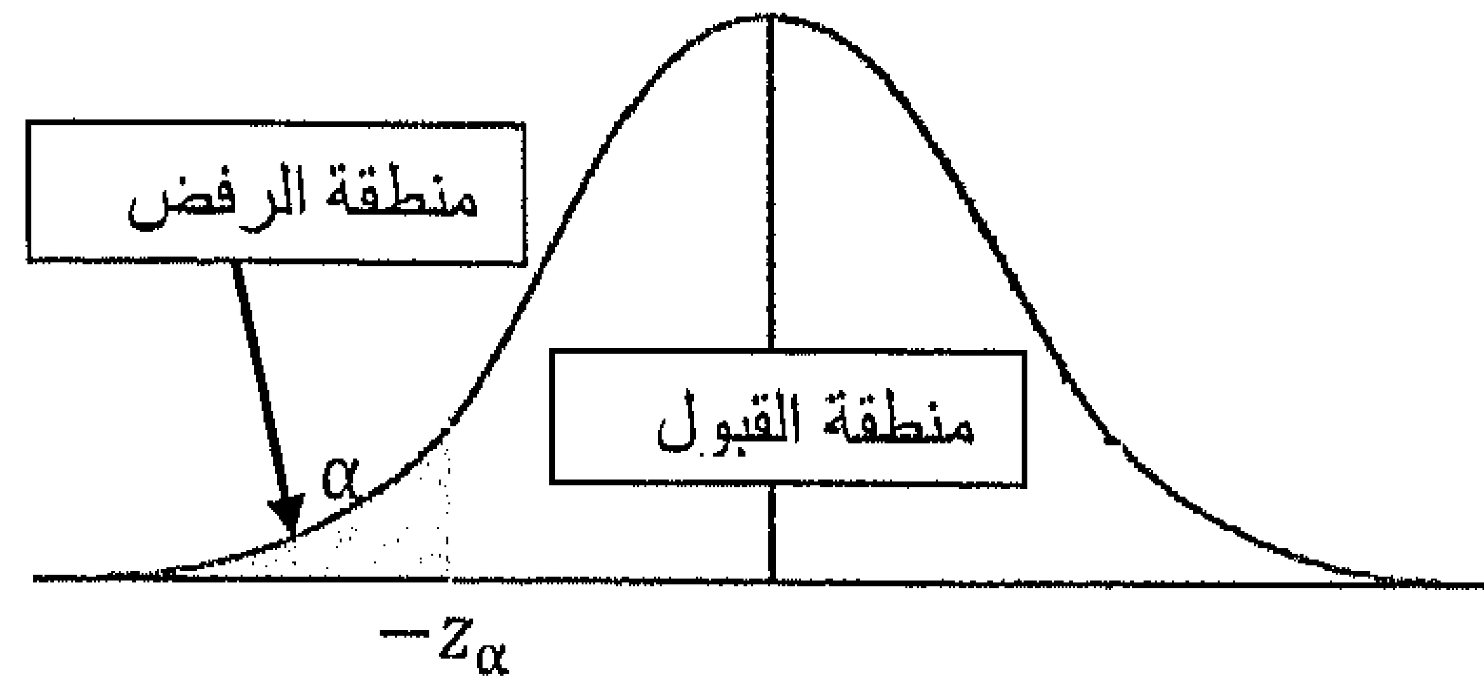
١- إذا كان الفرض البديل $H_1: \mu \neq \mu_0$ فإننا سوف نقوم بتقسيم قيمة مستوى المعنوية α إلى جزئيين متساويين $\alpha/2$ على أطراف المنحنى ومن ثم فيوجد منطقتا رفض ويسمى الاختبار في هذه الحالة بـ(اختبار ذو طرفين).



٢- إذا كان الفرض البديل $H_1: \mu > \mu_0$ فإنه سوف توجد منطقة رفض واحدة تقع على يمين المنحنى وفي هذه الحالة يسمى (الاختبار ذو طرف واحد- one tailed)



٣- إذا كان الفرض البديل $H_1: \mu < \mu_0$ فإنه سوف توجد منطقة رفض واحدة تقع على يسار المنحنى ويسمى (الاختبار ذو طرف واحد -one tailed)



ومن منحنى التوزيع الطبيعي القياسي فإن

$$\Phi(z_\alpha) = P[z \le z_\alpha] = 1 - \alpha,$$

$$\Phi(z_{\alpha/2}) = P[z \le z_{\alpha/2}] = 1 - \alpha/2$$

ويمكن تعيين حدود المناطق الحرجة $z_\alpha , z_{\alpha/2}$ من جدول التوزيع الطبيعي القياسي.

الحالة الثانية: إذا كان إحصاء الاختبار هو t سوف نستخدم منحنى توزيع t لتعيين حدود المناطق الحرجة ومن جدول توزيع t نوجد كلاً من القيم التالية والتي تمثل حدود المناطق الحرجة:

$$t_\alpha = t(\alpha, n-1),$$

$$t_{\alpha/2} = t(\alpha/2, n-1)$$

٥- القرار: سوف نرفض فرض العدم إذا كانت قيمة إحصاء الاختبار تقع في منطقة الرفض ونقبل فرض العدم إذا كانت قيمة إحصاء الاختبار تقع في منطقة القبول.

مثال (١١-١٧)

أخذت عينة عشوائية مكونة من 36 طالباً من طلاب الجامعة، فوجد أن متوسط طول الطالب في العينة هو 170 سم والمطلوب اختبار الفرض القائل بأن متوسط أطوال طلاب الجامعة يقل عن 175 سم إذا علم أن الانحراف المعياري لأطوال طلاب الجامعة هو 12 سم وذلك عند مستوى معنوية 0.01.

الحل

من المعطيات $n = 36,\ \sigma = 12,\ \bar{X} = 170,\ \mu_0 = 175$

١- الفروض الإحصائية

$$H_0: \mu = 175$$

$$H_1: \mu < 175$$

٢- مستوى المعنوية $\alpha = 0.01$

٣- إحصاء الاختبار: بما أن حجم العينة كبير n >30 فإن إحصاء الاختبار هو

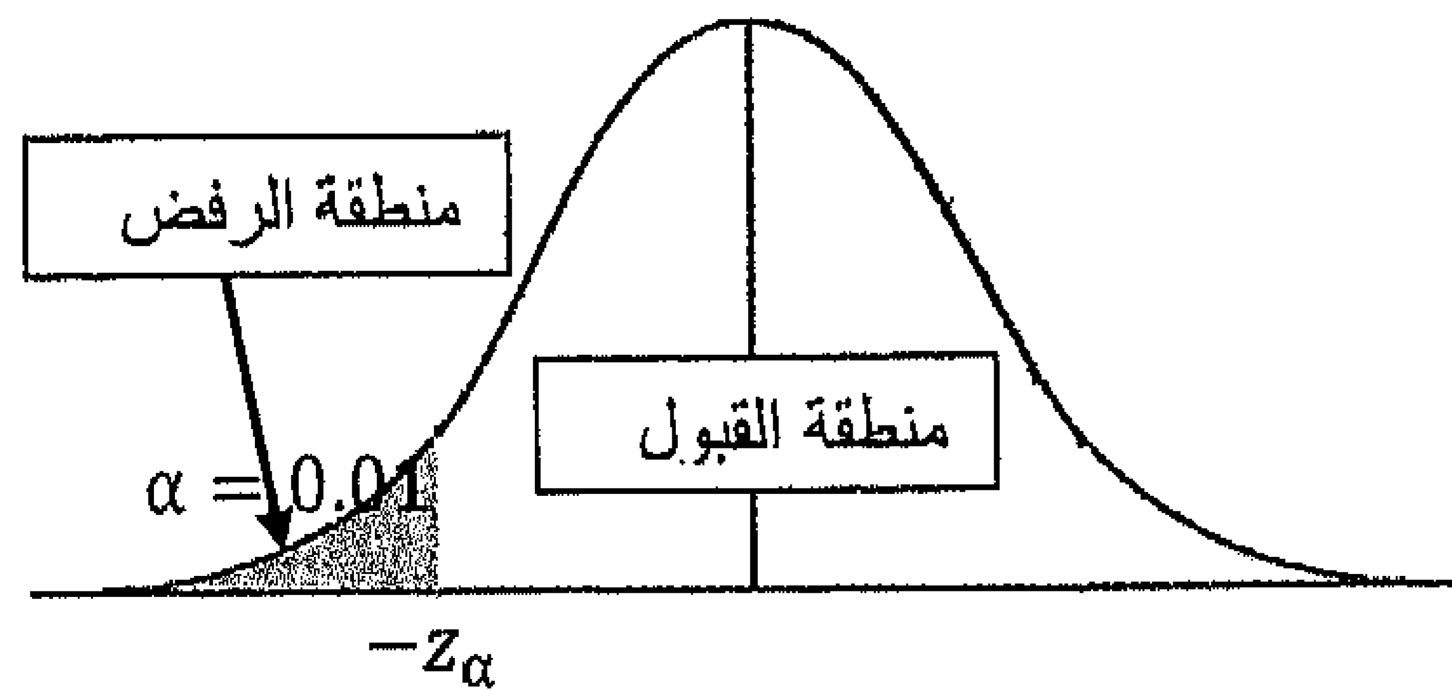
$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}}$$

وحيث إن حجم المجتمع غير محدود فإن

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{36}} = 2$$

$$\therefore Z = \frac{170 - 175}{2} = -2.5$$

٤- المناطق الحرجة: من منحنى التوزيع الطبيعي القياسي



$$\Phi(z_\alpha) = 1 - \alpha = 1 - 0.01 = 0.99$$

من جدول التوزيع الطبيعي القياسي فإن

$$z_\alpha = 2.33, \quad \Rightarrow -z_\alpha = -2.33$$

٥- القرار: سوف نرفض فرض العدم وذلك لأن إحصاء الاختبار يقع في منطقة الرفض، ونقبل الفرض البديل القائل بأن متوسط أطوال طلاب الجامعة يقل عن 175 سم.

مثال (١١-١٨)

أخذت عينة عشوائية مكونة من 64 طالبة من إحدى كليات البنات بجامعة سلمان بن عبدالعزيز والتي بها 1000 طالبة. وكان متوسط الطول لهن هو 162 سم. اختبر الفرض القائل بأن متوسط طول الطالبة بهذه الكلية أكبر من 160سم، عند مستوى معنوية 0.05 علماً بأن أطوال الطالبات بهذه الكلية يتبع التوزيع الطبيعي بانحراف معياري مقداره 4سم.

الحل

من المعطيات $n = 64, N = 1000, \sigma = 4, \bar{X} = 162, \mu_0 = 160$ فإن

١- الفروض الإحصائية

$$H_0: \mu = 160$$

$$H_1: \mu > 160$$

٢- مستوى المعنوية $\alpha = 0.05$

٣- إحصاء الاختبار: بما إن المجتمع يتبع التوزيع الطبيعي وتباينه معلوم فإن إحصاء الاختبار هو

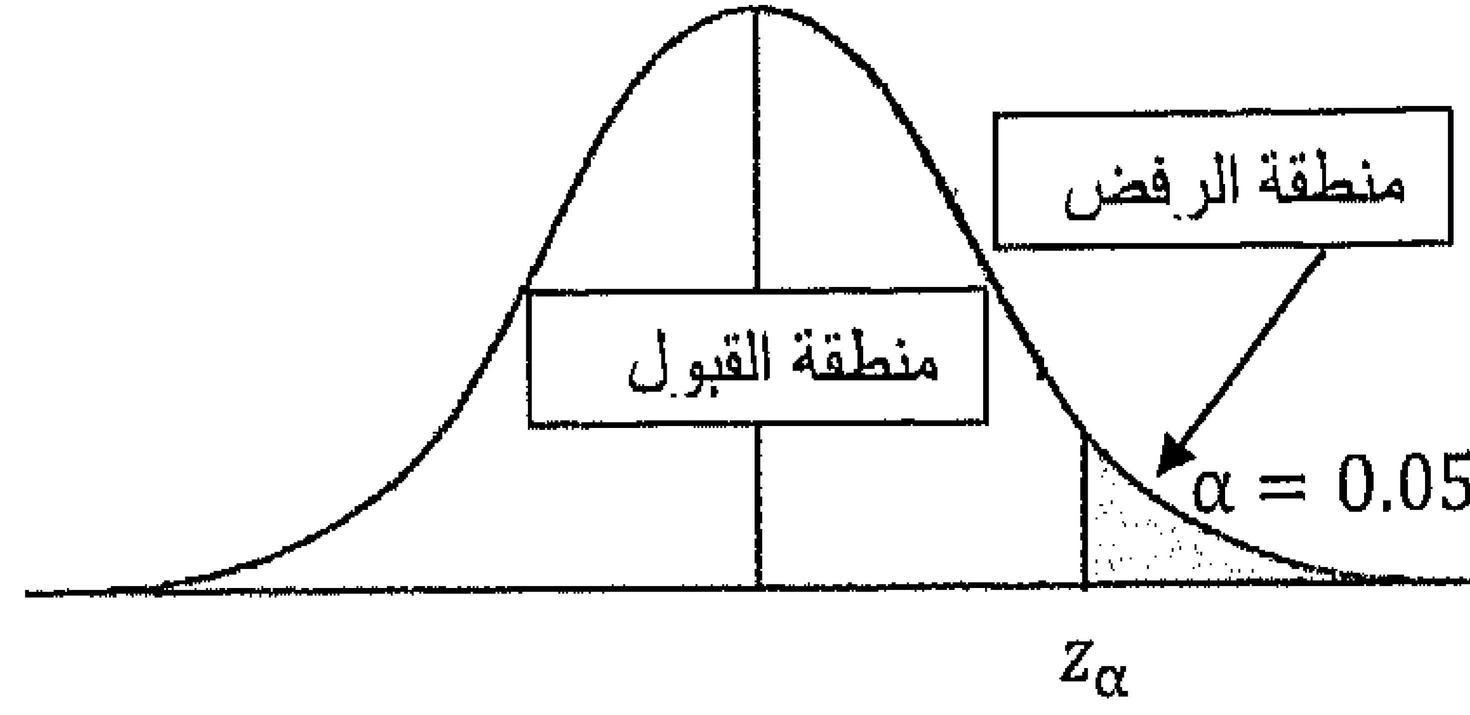
$$Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}}$$

وحيث إن المجتمع محدود فإن

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{4}{\sqrt{64}}\sqrt{\frac{1000-64}{1000-1}} = 0.48398$$

$$\therefore Z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{162 - 160}{0.48398} = 4.1324$$

٤- المناطق الحرجة: من منحنى التوزيع الطبيعي القياسي



$$\Phi(z_\alpha) = 1 - \alpha = 1 - 0.05 = 0.95$$

من جدول التوزيع الطبيعي القياسي فإن $z_\alpha = 1.645$

٥- القرار: سوف نرفض فرض العدم وذلك لأن إحصاء الاختبار يقع في منطقة الرفض. ونقبل الفرض البديل القائل بأن متوسط طول الطالبات بهذه الكلية أكبر من 160سم.

مثال(١١-١٩)

من بيانات عام 1983م تبين أن متوسط أعمار العاملين بإحدى الهيئات الحكومية هو 38 سنة وفي عام 2004م تم اختيار عينة من العاملين بهذه الهيئة حجمها 25 شخصا فكانت أعمارهم كما يلي:

28	34	43	45	45	33	45	26	50	46	39	41	39
44	39	40	41	31	43	35	40	44	46	33	37	

هل تدل هذه المشاهدات على أن متوسط أعمار العاملين في عام 2004م قد اختلف عن متوسط أعمارهم في عام 1983م وذلك عند مستوى معنوية 0.05؟ علماً بأن أعمار العاملين بتلك الهيئة يتبع التوزيع الطبيعي.

الحل

من المعطيات $n = 25, \mu_0 = 38$ فإن

١- الفروض الإحصائية

$$H_0: \mu = 38$$

$$H_1: \mu \neq 38$$

٢- مستوى المعنوية $\alpha = 0.05$

٣- إحصاء الاختبار: بما أن المجتمع يتبع التوزيع الطبيعي وتباينه غير معلوم، وحجم العينة صغير فإن إحصاء الاختبار هو

$$T = \frac{\bar{X} - \mu_0}{s_{\bar{x}}}$$

من البيانات المعطاة فإن $\sum x = 987, \sum x^2 = 39851$ فإنه يمكن تعيين الوسط الحسابي والتباين للبيانات كما يلي:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{987}{25} = 39.48$$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right]$$

$$= \frac{1}{24}[\,39851 - 25(39.48)^2] = 36.84333$$

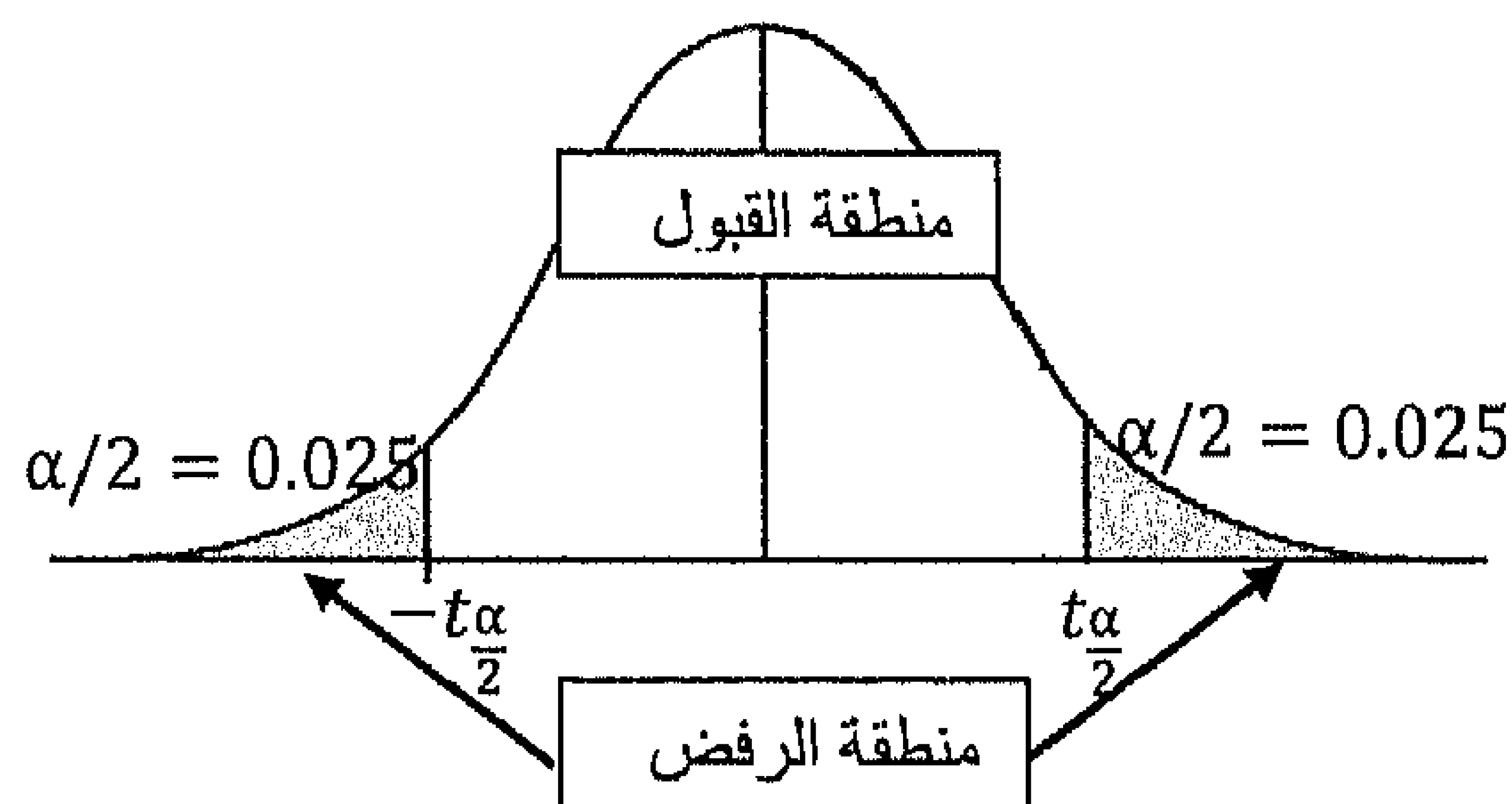
$$\Longrightarrow s = 6.06987$$

وحيث إن المجتمع غير محدود فإن

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{6.06987}{\sqrt{25}} = 1.21397$$

$$\therefore T = \frac{39.48 - 38}{1.21397} = 1.21914$$

٤- المناطق الحرجة: من منحنى توزيع t فان



من جدول توزيع t فإن

$$t_{\frac{\alpha}{2}} = t\left(n-1, \frac{\alpha}{2}\right) = t(24, 0.025) = 2.069,$$

$$-t_{\alpha/2} = -2.069$$

٥- القرار: سوف نقبل فرض العدم وذلك لأن إحصاء الاختبار يقع في منطقة القبول. وهو أن متوسط أعمار العاملين في عام 2004م لا يختلف عن متوسط أعمارهم في عام 1983م

مثال(١١-٢٠)

إذا كانت أعمار طلاب احدي الجامعات تتبع التوزيع الطبيعي بانحراف معياري مقداره 3 سنوات. ادعى أحد الباحثين الاجتماعيين بأن متوسط عمر الطالب في هذه الجامعة هو 24 سنة، ولاختبار هذا الادعاء اختيرت عينة من 20 طالباً فكان متوسط أعمارهم هو 23 سنة. فهل معنى ذلك أن متوسط عمر طلاب الجامعة يقل عن 24 سنة؟ استخدم مستوى معنوية 0.05.

الحل

من المعطيات $n = 20, \sigma = 3, \bar{X} = 23, \mu_0 = 24$ فإن

١- الفروض الإحصائية

$$H_0: \mu = 24$$

$$H_1: \mu < 24$$

٢- مستوى المعنوية $\alpha = 0.01$

٣- إحصاء الاختبار: بما أن المجتمع يتبع التوزيع الطبيعي والتباين معلوم فإن إحصاء الاختبار هو

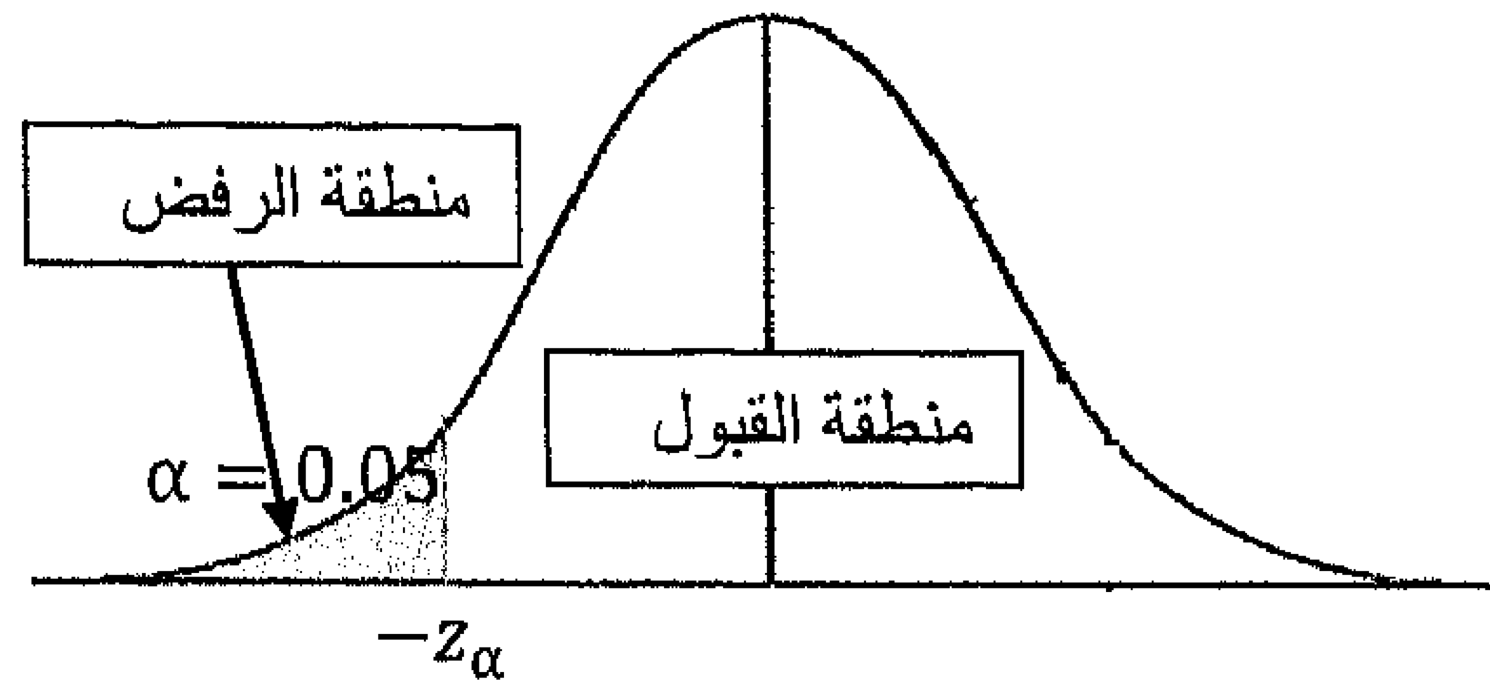
$$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}}$$

وحيث إن المجتمع غير محدود فإن

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{20}} = 0.6708$$

$$\therefore z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{23 - 24}{0.6708} = -1.4907$$

٤- المناطق الحرجة: من منحنى التوزيع الطبيعي القياسي فإن



$$\Phi(z_\alpha) = 1 - \alpha = 1 - 0.05 = 0.95$$

من جدول التوزيع الطبيعي القياسي فإن

$$\Rightarrow z_\alpha = 1.645, -z_\alpha = -1.645$$

٥- القرار: سوف نقبل فرض العدم وذلك لأن إحصاء الاختبار يقع في منطقة القبول. وبالتالي فإن متوسط عمر طلاب الجامعة لا يختلف عن 24 سنة.

مثال(١١-٢١)

إذا كانت أسعار إحدى السلع تتبع توزيعاً طبيعياً وكان متوسط سعر السلعة عام 1987م هو 38.5 ريالاً للكيلوجرام. في عام 1995م تم اختيار عينة من أسعار هذه السلعة حجمها 15 فكان وسطها الحسابي هو 40 ريالاً والانحراف المعياري هو 4 ريالات. هل تدل هذه البيانات على اختلاف متوسط سعر السلعة في العامين؟ استخدم مستوى معنوية 0.01

الحل

من المعطيات $\mu_0 = 38.5\,, n = 15, \bar{x} = 40, s = 4$ فإن

١- الفروض الإحصائية

$$H_0: \mu = 38.5$$

$$H_1: \mu \neq 38.5$$

٢- مستوى المعنوية $\alpha = 0.01$

٣- إحصاء الاختبار: بما أن المجتمع يتبع التوزيع الطبيعي وتباينه غير معلوم، وحجم العينة صغير فإن إحصاء الاختبار هو

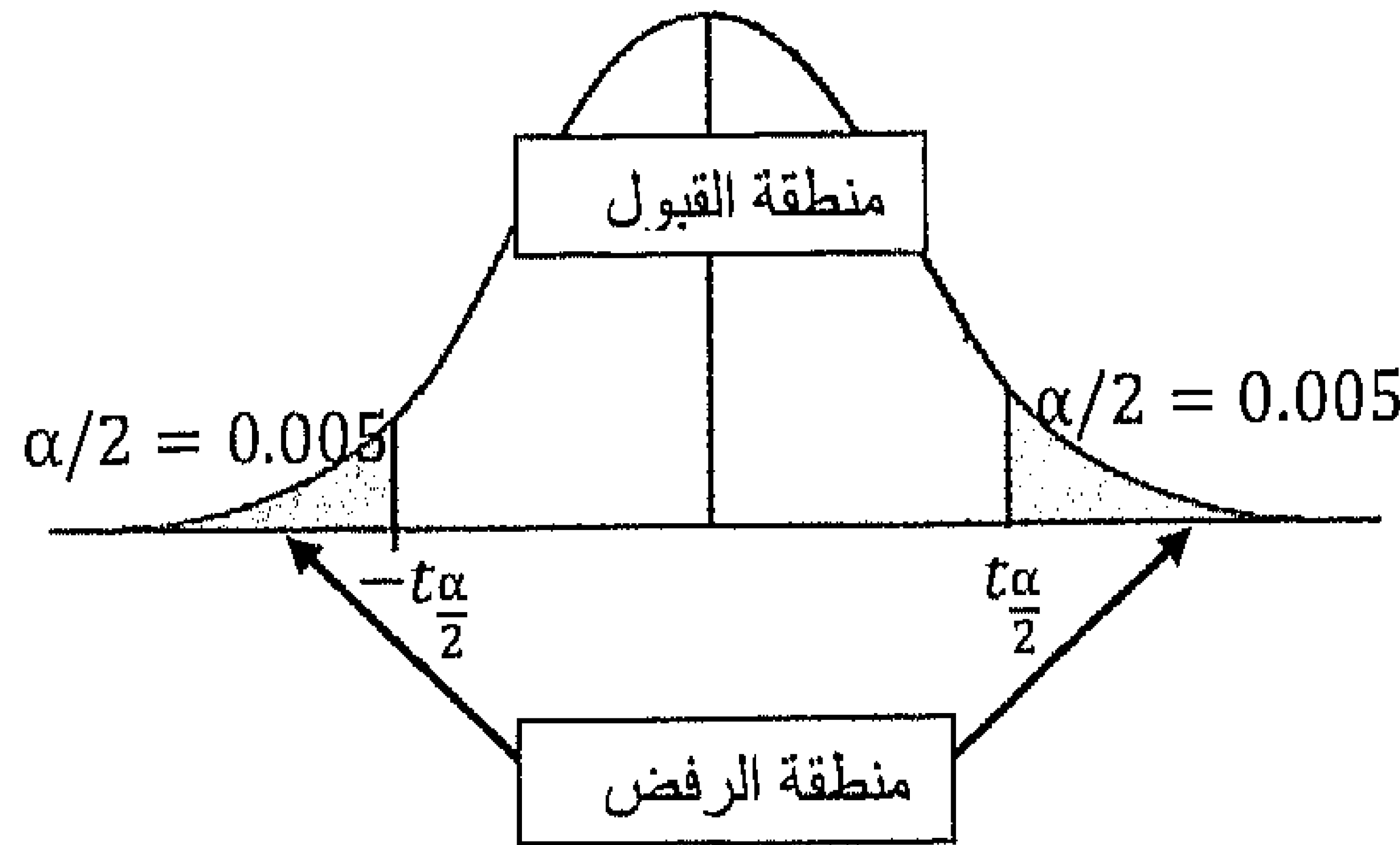
$$T = \frac{\bar{X} - \mu_0}{s_{\bar{x}}}$$

وحيث إن المجتمع غير محدود فإن

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{4}{\sqrt{15}} = 1.0328$$

$$\therefore T = \frac{40 - 38.5}{1.0328} = 1.4524$$

٤- المناطق الحرجة:من منحنى توزيع t فإن



من جدول توزيع t فإن

$$t_{\frac{\alpha}{2}} = t\left(n - 1, \frac{\alpha}{2}\right) = t(14, 0.005) = 2.977,$$

$$-t_{\alpha/2} = -2.977$$

٥- القرار: سوف نقبل فرض العدم وذلك لأن إحصاء الاختبار يقع في منطقة القبول. وبالتالي فإنه لا يوجد اختلاف في متوسط سعر السلعة في العامين.

١١-٣-٢ اختبار فرض حول نسبة توفر صفة معينة في المجتمع للعينات الكبيرة

إذا كانت p هى نسبة توفر صفة معينة في المجتمع، واخترنا عينات كبيرة من هذا المجتمع، فمن توزيعات المعاينة فإن نسبة توفر نفس الصفة في العينة r سوف تتبع توزيعاً طبيعياً وسطه p وتباينه $\frac{p(1-p)}{n}$.

وبفرض أننا نريد اختبار ادعاء حول نسبة وجود صفة ما في المجتمع فسوف نتبع الخطوات التالية:

١- الفروض الإحصائية

$$H_0: p = p_0$$

$$H_1: \begin{cases} p \neq p_0 \\ p > p_0 \\ p < p_0 \end{cases}$$

٢- مستوى المعنوية α

٣- إحصاء الاختبار هو

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} \sim N(0,1)$$

حيث

$$\sigma_{\hat{p}}^2 = \begin{cases} \dfrac{p_0(1-p_0)}{n}, & \text{اذا كان المجتمع غير محدود} \\ \dfrac{p_0(1-p_0)}{n}\left(\dfrac{N-n}{N-1}\right), & \text{اذا كان المجتمع محدود} \end{cases}$$

وحيث إن إحصاء الاختبار z يتبع التوزيع الطبيعي القياسي فإن المناطق الحرجة سوف تحدد كما سبق مع الوسط وكذلك عملية الرفض والقبول.

مثال(١١-٢٢)

أخذت عينة مكونة من 200 فرداً من سكان مدينة ما بها 14000 فرداً فوجد أن عدد المدخنين منهم هو 80 مدخن. اختبر الفرض القائل بأن نسبة المدخنين بهذه المدينة هو 24% وذلك عند مستوى معنوية 0.01

الحل

بفرض أن عدد المدخنين في العينة هو x ومن المعطيات $n = 200, x = 80, N = 14000$ وبالتالي فإن

$p_0 = \frac{24}{100} = 0.24,$

$\hat{p} = r = \frac{80}{200} = 0.4$

١- الفروض الإحصائية

$$H_0: p = 0.24$$
$$H_1: p \neq 0.24$$

٢- مستوى المعنوية $\alpha = 0.01$

٣- إحصاء الاختبار: بما أن حجم العينة كبير فإن إحصاء الاختبار هو

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$$

حيث إن المجتمع محدود فإن

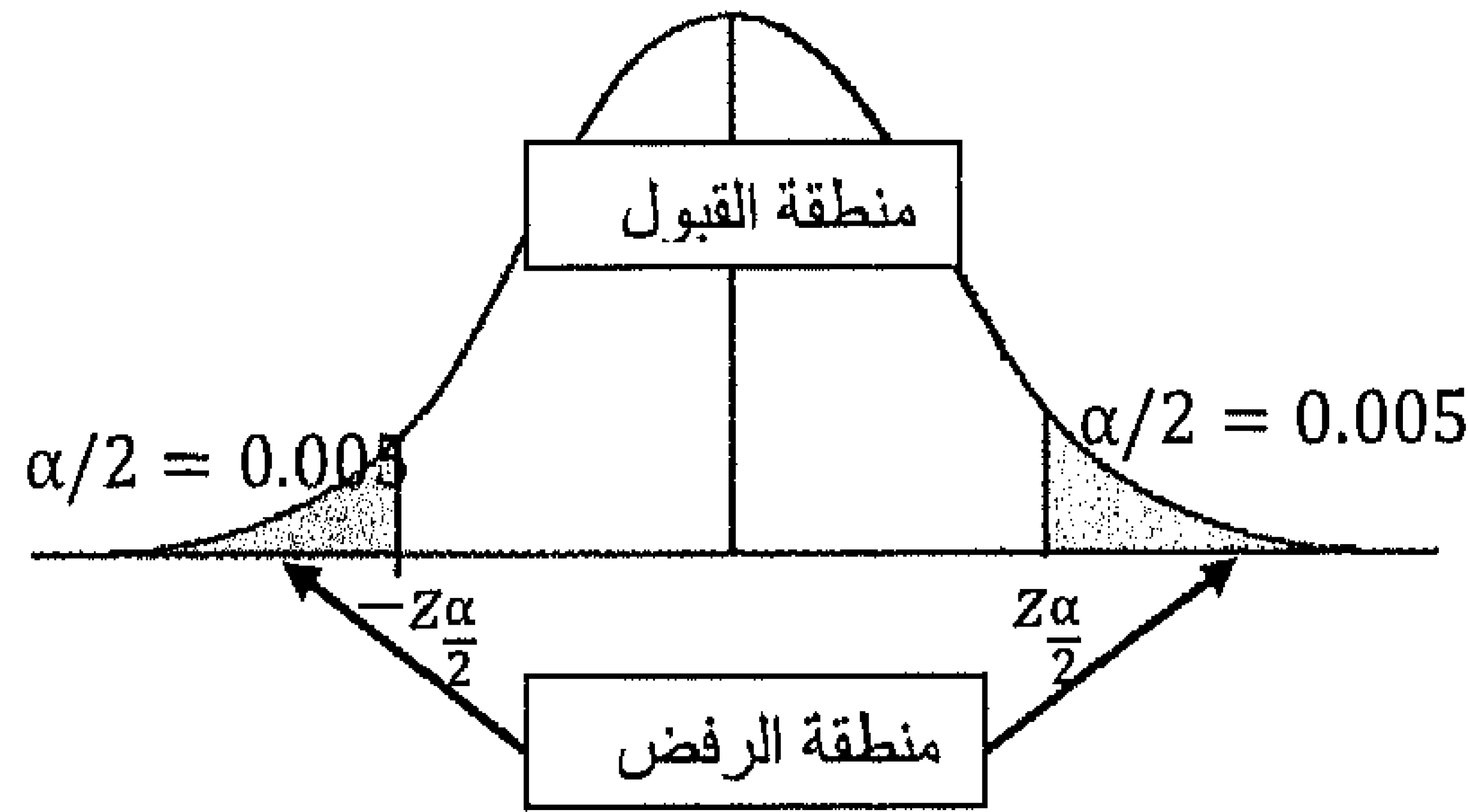
$$\sigma_{\hat{p}}^2 = \frac{p_0(1-p_0)}{n}\left(\frac{N-n}{N-1}\right)$$

$$= \frac{0.24(1-0.24)}{200}\left(\frac{14000-200}{14000-1}\right) = 0.000899$$

$$\Longrightarrow \sigma_{\hat{p}} = 0.02998$$

$$\therefore z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.4 - 0.24}{0.02998} = 5.33689$$

٤- المناطق الحرجة: من منحنى التوزيع الطبيعي القياسي والفرض البديل فإن



$$\Phi(z_{\alpha/2}) = 1 - \alpha/2 = 1 - 0.005 = 0.995$$

من جدول التوزيع الطبيعي القياسي فإن

$$\Longrightarrow z_{\alpha/2} = 2.575, -z_{\alpha/2} = -2.575$$

٥- القرار: سوف نرفض فرض العدم وذلك لأن إحصاء الاختبار يقع في منطقة الرفض. وبالتالي فإن نسبة المدخنين بهذه المدينة يختلف عن 24%

مثال(١١-٢٣)

أخذت عينة مكونة من 250 نباتاً من نبات القمح تبين أن 59 نباتاً تتوفر فيها صفة معينة. وطبقاً لقانون مندل للوراثة فإن 25% من محصول القمح يجب أن تتوفر فيه هذه الصفة. فهل تتفق هذه النتيجة مع قانون مندل للوراثة؟ استخدم مستوى معنوية 0.05.

الحل

بفرض أن عدد النباتات التي تتوفر بها الصفة في العينة هو x ومن المعطيات,$n = 250, x = 59$ فإن

$$p_0 = \frac{25}{100} = 0.25,$$

$$\hat{p} = \frac{59}{250} = 0.236$$

١- الفروض الإحصائية

$$H_0: p = 0.25$$

$$H_1: p \neq 0.25$$

٢- مستوى المعنوية $\alpha = 0.05$

٣- إحصاء الاختبار: بما أن حجم العينة كبير فإن إحصاء الاختبار هو

$$z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}}$$

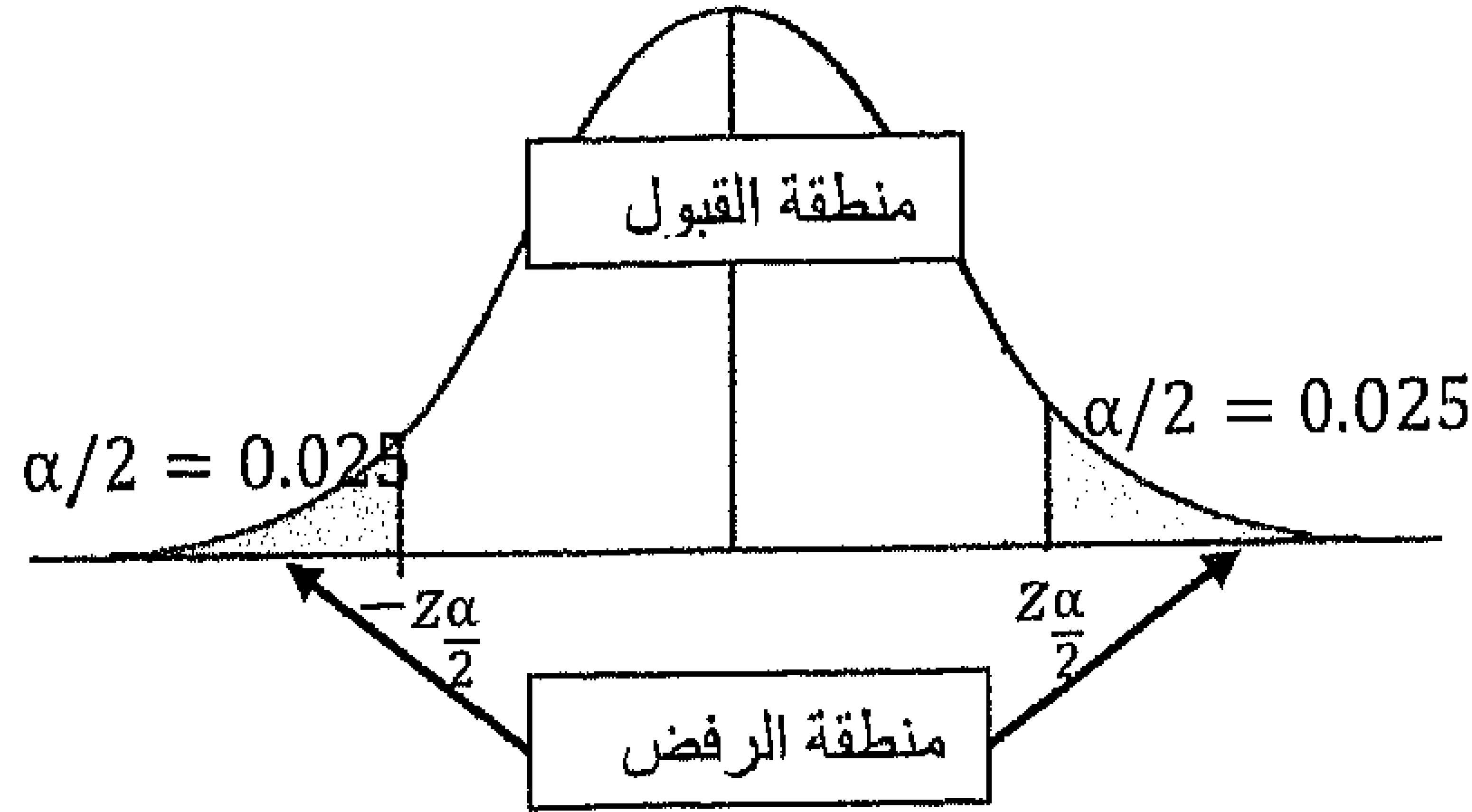
حيث إن المجتمع غير محدود فإن

$$\sigma_{\hat{p}}^2 = \frac{p_0(1 - p_0)}{n} = \frac{0.25(1 - 0.25)}{250} = 0.00075$$

$$\Longrightarrow \sigma_{\hat{p}} = 0.02739$$

$$\therefore z = \frac{\hat{p} - p_0}{\sigma_{\hat{p}}} = \frac{0.236 - 0.25}{0.02739} = -0.5111$$

٤- المناطق الحرجة: من منحنى التوزيع الطبيعي القياسي والفرض البديل فإن



$$\Phi(z_{\alpha/2}) = 1 - \alpha/2 = 1 - 0.025 = 0.975$$

من جدول التوزيع الطبيعي القياسي فإن

$$\Longrightarrow z_{\alpha/2} = 1.96, \quad -z_{\alpha/2} = -1.96$$

٥- القرار: سوف نقبل فرض العدم وذلك لأن إحصاء الاختبار يقع في منطقة القبول. وبالتالي فإنه تتفق هذه النتيجة مع قانون مندل للوراثة.

١١-٣-٣ اختبار فرض حول تباين المجتمع

من معالم المجتمع التي تؤدي دوراً مهما هو التباين σ^2 وقد نحتاج أحياناً لدراسة صحة ادعاء حول التباين أو الانحراف المعياري لمجتمع يتبع التوزيع الطبيعي. ولذلك سوف نقوم بسحب عينة عشوائية من المجتمع $N(\mu, \sigma^2)$ وبتعيين التباين لتلك العينة فإنه يمكن اختبار الادعاء بأن تباين المجتمع σ^2 يساوى أو يختلف أو أكبر أو أقل من قيمة معينة σ_0^2 باتباع الخطوات التالية:

١- الفروض الإحصائية

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \begin{cases} \sigma^2 \neq \sigma_0^2 \\ \sigma^2 > \sigma_0^2 \\ \sigma^2 < \sigma_0^2 \end{cases}$$

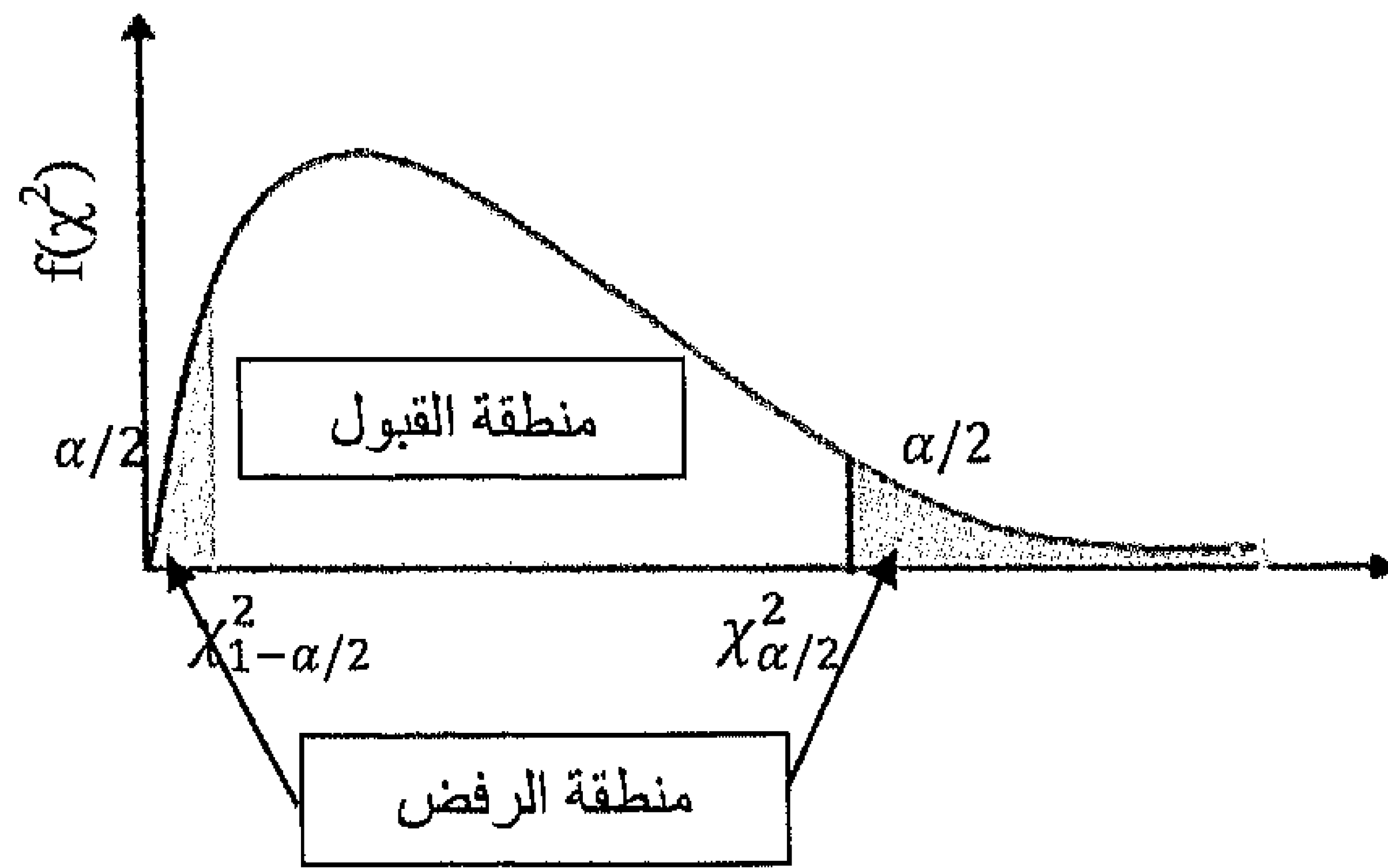
٢- مستوى المعنوية α

٣- إحصاء الاختبار هو

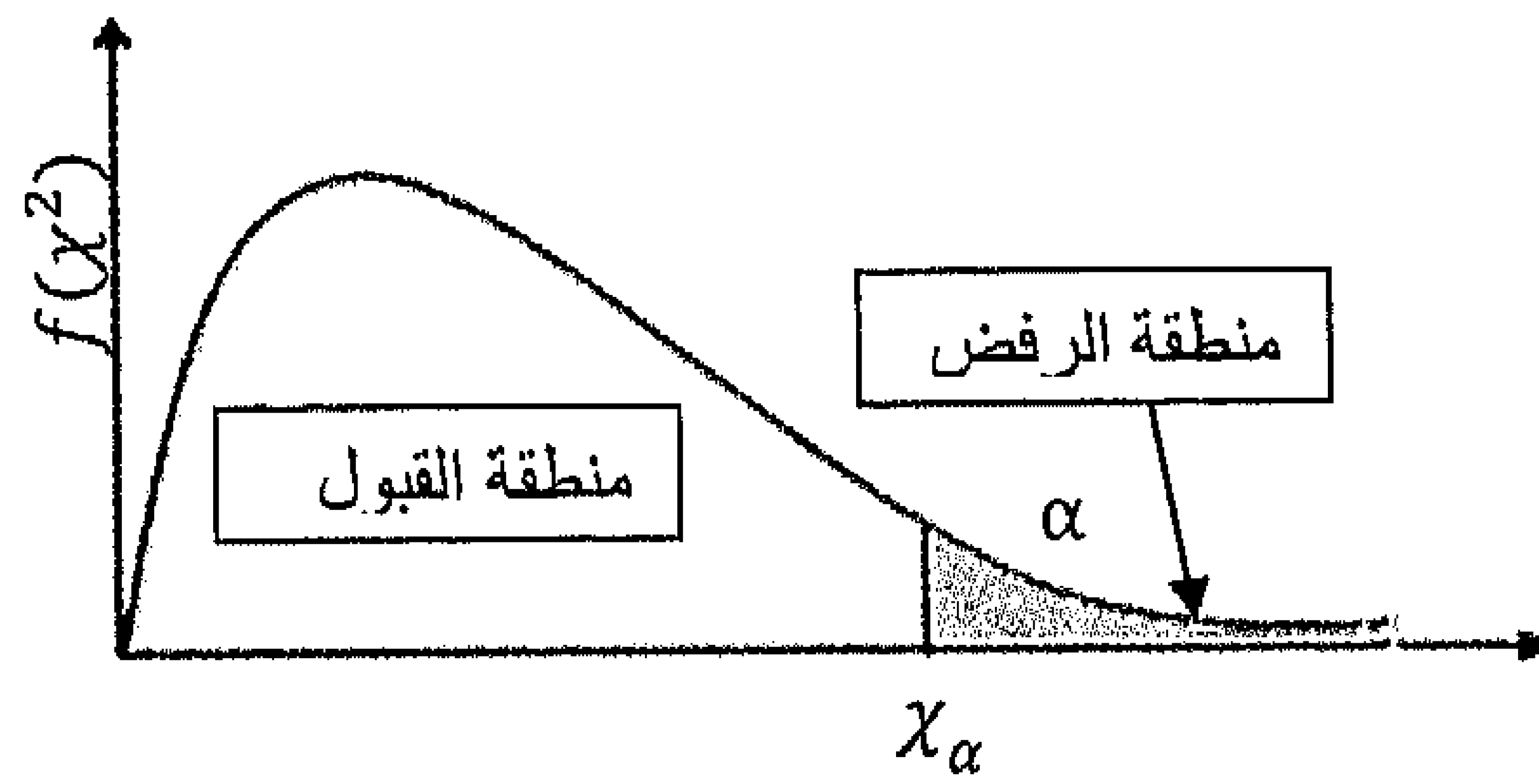
$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2(n-1)$$

٤- تحديد المناطق الحرجة: حيث إن إحصاء الاختبار يتبع توزيع مربع كاي بدرجة حرية $n-1$ فإننا سوف نستخدم منحنى توزيع مربع كاي لتحديد المناطق الحرجة وبالتالي سوف نستخدم جدول توزيع مربع كاي لتعيين حدود المناطق الحرجة كالتالي:

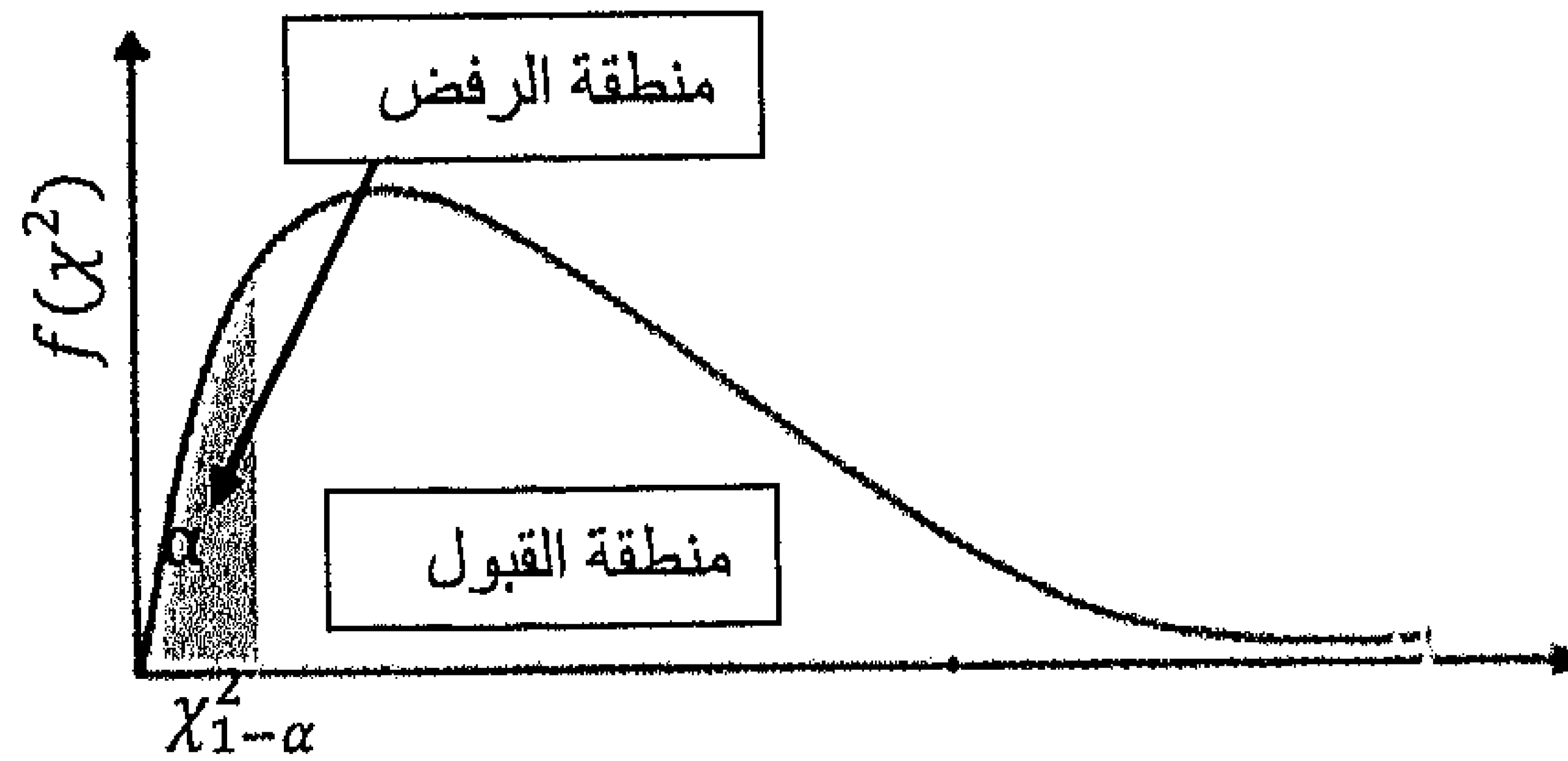
- إذا كان الفرض البديل $H_1: \sigma^2 \neq \sigma_0^2$



- إذا كان الفرض البديل $H_1: \sigma^2 > \sigma_0^2$



- إذا كان الفرض البديل $H_1: \sigma^2 < \sigma_0^2$



ومن منحنى توزيع مربع كاي فإن

$$\chi^2_{\alpha} = \chi^2(n-1,\alpha),$$

$$\chi^2_{1-\alpha} = \chi^2(n-1,1-\alpha)$$

$$\chi^2_{\alpha/2} = \chi^2(n-1,\alpha/2),$$

$$\chi^2_{1-\alpha/2} = \chi^2(n-1,1-\alpha/2)$$

ومن جدول توزيع مربع كاي يمكن تعيين كل من القيم السابقة والتي تمثل حدود المناطق الحرجة.

٥- القرار: سوف نرفض فرض العدم إذا كان قيمة إحصاء الاختبار تقع في منطقة الرفض ونقبل فرض العدم إذا كانت قيمة إحصاء الاختبار تقع في منطقة القبول. وذلك بمقارنة قيمة إحصاء الاختبار بحدود الفترات.

مثال(١١-٢٤)

لاختبار الفرض بأن تباين المجتمع $\sigma^2 = 127$ اختيرت عينة عشوائية حجمها 16 مفردة فوجد أن تباينها هو 112.07 استخدم مستوى معنوية 0.02 لاختبار هذا الفرض علماً بأن المجتمع يتبع التوزيع الطبيعي.

الحل

من المعطيات فإن $\sigma_0^2 = 127, n = 16, s^2 = 112.07$ وبالتالي فإن:

١- الفروض الإحصائية

$$H_0: \sigma^2 = 127$$

$$H_1: \sigma^2 \neq 127$$

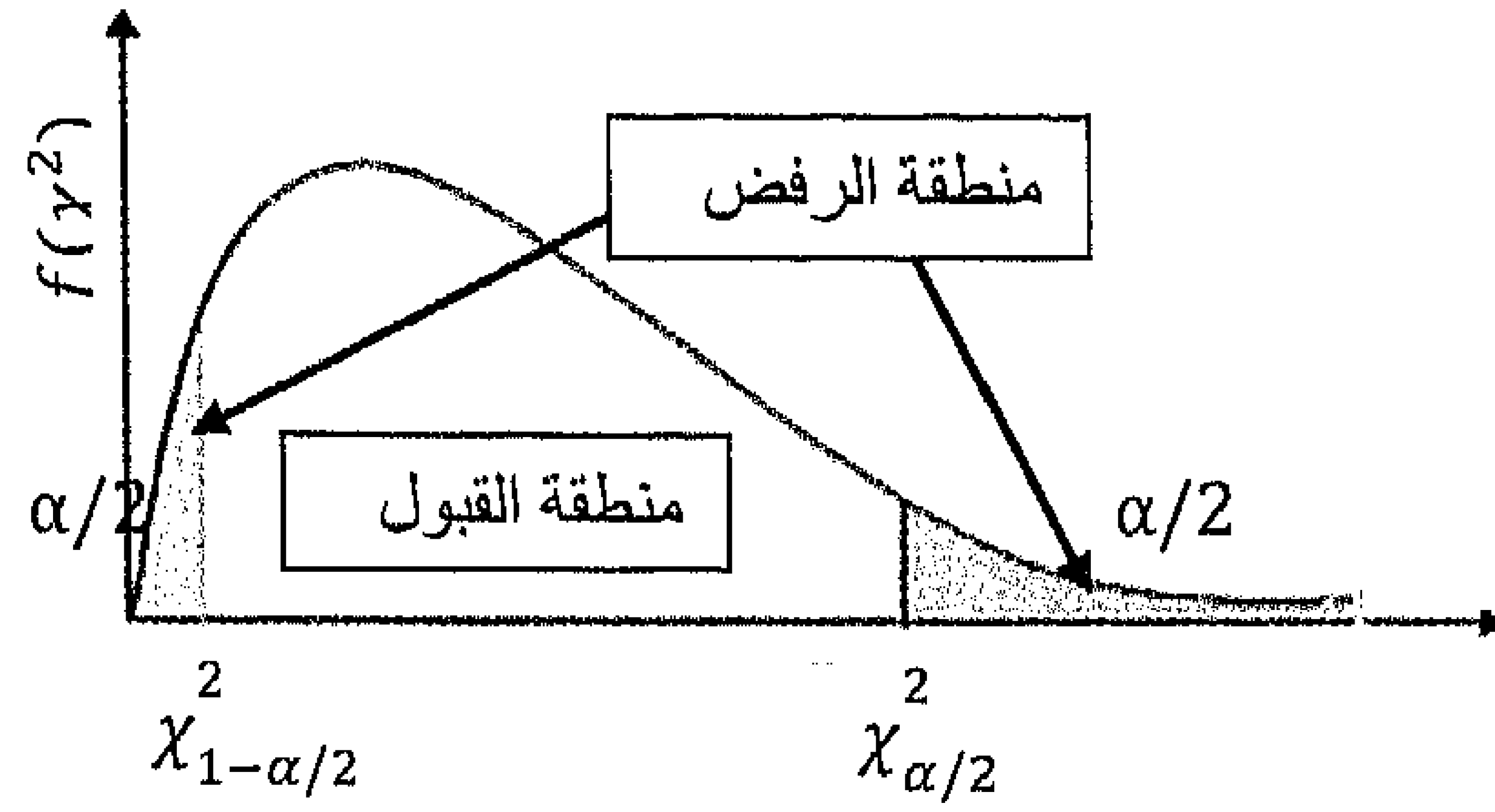
٢- مستوى المعنوية $\alpha=0.02$

٣- إحصاء الاختبار هو

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$= \frac{(16-1)\times(112.07)}{127} = 13.2366$$

٤- تحديد المناطق الحرجة: حيث إن إحصاء الاختبار تتبع توزيع مربع كاي بدرجة حرية $n-1$ فإننا سوف نستخدم منحنى توزيع مربع كاي لتحديد المناطق الحرجة كما يلي:



$$\alpha = 0.02 \Rightarrow \frac{\alpha}{2} = 0.01$$

ومن جدول توزيع مربع كاي فإن

$$\chi^2_{\alpha/2} = \chi^2(n-1, \alpha/2) = \chi^2(15, 0.01) = 30.578,$$

$$\chi^2_{1-\alpha/2} = \chi^2(n-1, 1-\alpha/2) = \chi^2(15, 0.99) = 5.229$$

٥- القرار: سوف نقبل فرض العدم لأن قيمة إحصاء الاختبار تقع في منطقة القبول. ومن ثم فإن تباين المجتمع لا يختلف عن 127.

مثال (١١-٢٥)

إذا كان من المعروف أن الضغط الكلى الداخلي لأسطوانات الغاز المعبأة بواسطة إحدى شركات تعبئة الغاز الطبيعي تتبع التوزيع الطبيعي بمتوسط 120 رطلاً لكل بوصة مربعة وتباين 0.25 وقد بدأت هذه الشركة في اتباع طريقة جديدة لتعبئة تلك الأسطوانات تحافظ على المتوسط عند 120 رطلاً ولكن على أمل أن تقلل من قيمة التباين. وبعد بدء التعبئة بالطريقة الجديدة اختيرت عينة من الأسطوانات حجمها 25 فكان الضغط لكل منها هو

119.99	119.98	120.68	119.31	120.20	120.75	120.12	120.57
119.40	120.29	120.30	119.96	120.36	119.27	119.76	119.94
119.82	119.91	120.60	119.71	119.59	120.04	119.47	119.46
119.99							

فهل تدل هذه البيانات على أن أمل الشركة قد تحقق باتباع الطريقة الجديدة للتعبئة؟ استخدم مستوي معنوية 0.05

الحل

من المعطيات $n = 25, \mu = 120, \sigma_0^2 = 0.25$ فإن

١- الفروض الإحصائية

$$H_0: \sigma^2 = 0.25$$

$$H_1: \sigma^2 < 0.25$$

٢- مستوى المعنوية $\alpha=0.05$

٣- إحصاء الاختبار هو

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

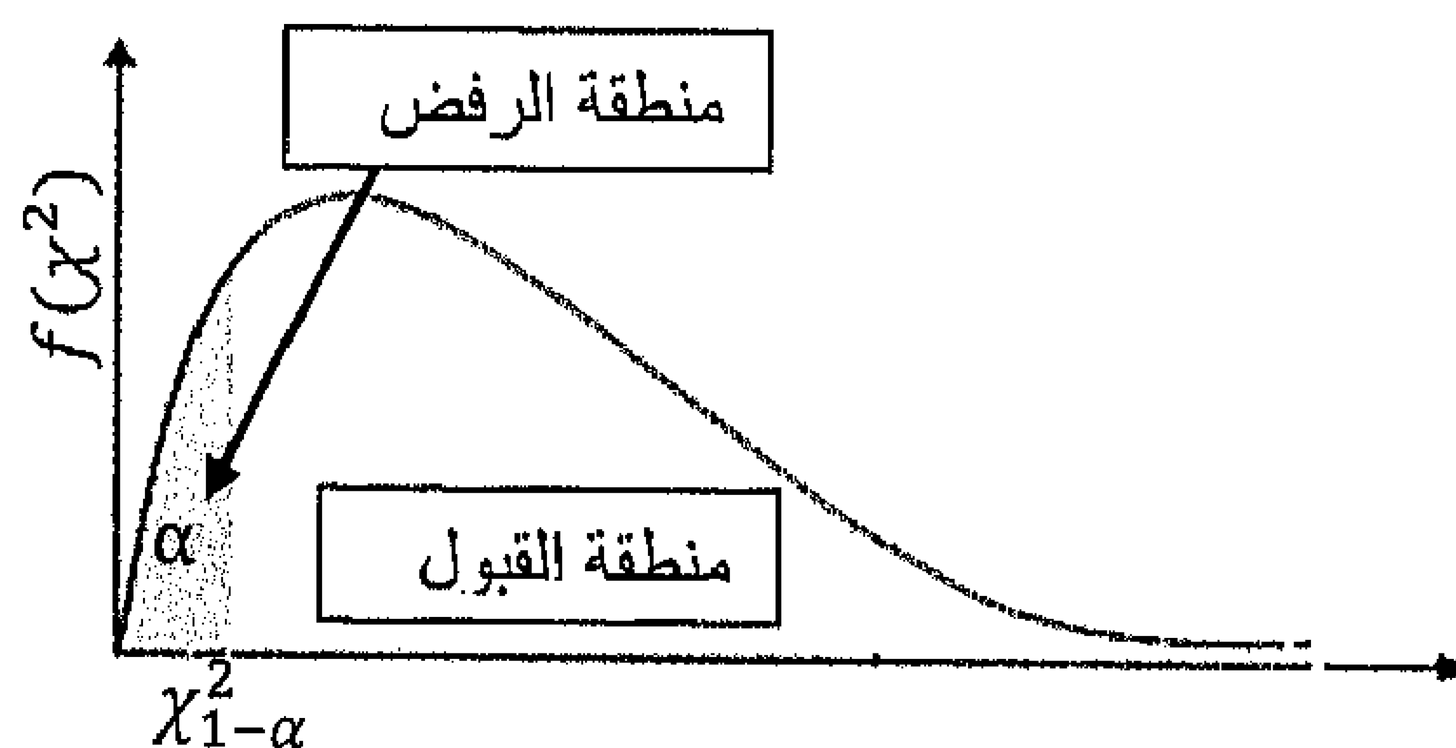
من بيانات العينة المعطاة فإن $\sum x = 2999.47, \sum x^2 = 359877.1671$ ومن ثم فإنه يمكن تعيين قيمة التباين كما يلي:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2\right]$$

$$= \frac{1}{25-1}\left[359877.1671 - 25\left(\frac{2999.47}{25}\right)^2\right] = 0.1815$$

$$\therefore \chi_0^2 = \frac{(25-1)(0.1815)}{0.25} = 17.424$$

٤- المناطق الحرجة: حيث إن إحصاء الاختبار تتبع توزيع مربع كاي بدرجة حرية $n-1$ فإننا سوف نستخدم منحنى توزيع مربع كاي لتحديد المناطق الحرجة



ومن جدول توزيع مربع كاي فإن

$$\chi_{1-\alpha}^2 = \chi^2(n-1, 1-\alpha) = \chi^2(24, 0.95) = 13.848$$

٥- القرار: سوف نقبل فرض العدم لأن إحصاء الاختبار تقع في منطقة القبول. ومن ثم فإن الضغط الداخلي لأسطوانات الغاز لم يتغير مع اتباع الطريقة الجديدة في التعبئة.

١١-٤ العلاقة بين فترات الثقة واختبارات الفروض:

توجد علاقة بين فترات الثقة واختبارات الفروض الإحصائية، وذلك بأنه يمكن إجراء عملية القبول أو الرفض لفرض العدم عن طريق تعيين حدود فترة الثقة للمعلم المجهول θ ثم مقارنة القيمة θ_0 بحدود الفترة فإذا كانت تقع في الفترة سوف نقبل فرض العدم وإذا كانت تقع خارج فترة الثقة سوف نرفض فرض العدم ونقبل الفرض البديل.

١١-٥ تطبيقات باستخدام برنامج SPSS

سوف نقوم بإجراء بعض التطبيقات باستخدام برنامج SPSS ولكننا نجد أن البرنامج يقدم لنا حالة واحدة فقط وهى الحالة التي يتبع فيها المجتمع التوزيع الطبيعي وتباينه غير معلوم وحجم العينة صغير والتي يكون فيها توزيع المعاينة هو توزيع t في كل من التقدير واختبارات الفروض الإحصائية حول الوسط لكن باقي الحالات الخاصة بالمتوسط μ فهي غير موجودة في برنامج SPSS وأيضا التقدير واختبار فرض حول النسبة والتباين فغير موجودة في البرنامج ولكنها موجودة جميعاً في برنامج Minitab وهذا ما سنقوم به في الجزء الثاني من هذا الكتاب وهو استخدام كل من SPSS, Minitab لإجراء التطبيقات الإحصائية لموضوع الاستدلال الإحصائي والذي سنتحدث عنه بالتفصيل ولكننا هنا سوف نكتفي بما هو متاح في برنامج SPSS.

تطبيق(١١-١):

فى مثال(١١-٩): أخذت عينة عشوائية مكونه من 9 أفدنة مزروعة بنوع معين من القمح فوجد أن إنتاجها بالإردب كالاتى 12 ,10 ,11 ,9 ,8 ,10.5 ,9.5 ,11.5 ,8.5 فاحسب 95% فترة ثقة لمتوسط إنتاجية الفدان من ذلك النوع من القمح، وذلك بفرض أن إنتاجية الفدان من ذلك النوع من القمح لها التوزيع الطبيعى.

الحل:

من معطيات المثال فإن المجتمع يتبع التوزيع الطبيعي وتباينه معلوم وحجم العينة $n = 9$ صغير لذا فإن $100\%(-1)$ فترة ثقة للمتوسط μ هي

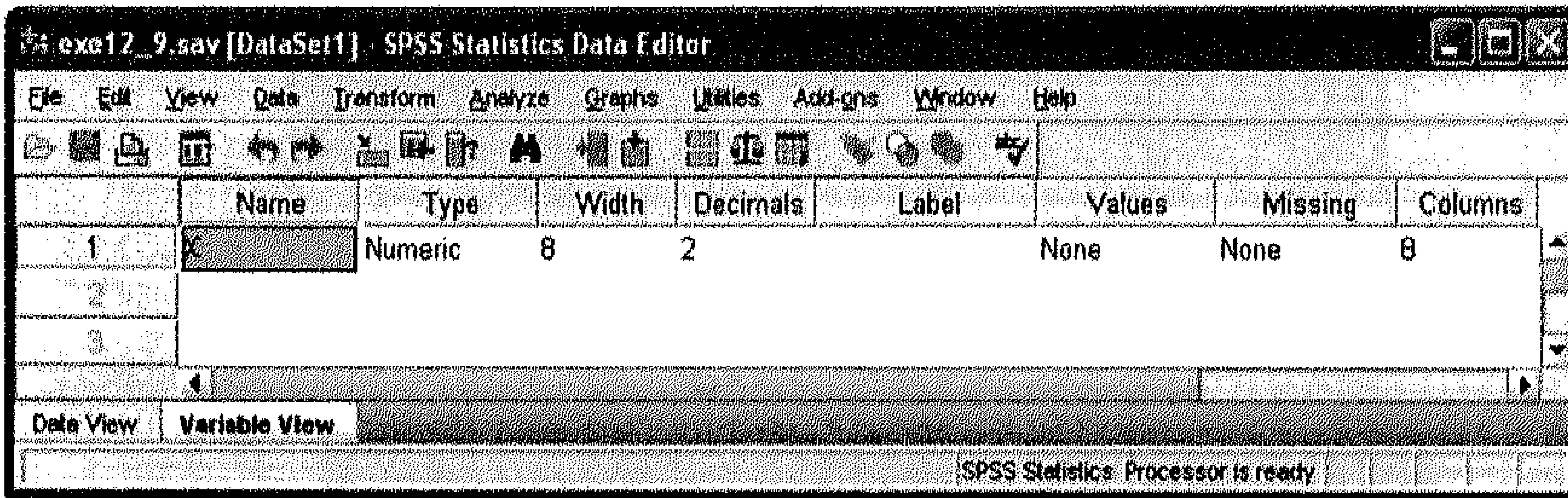
$$\bar{x} - t_{\frac{\alpha}{2}}\, s_{\bar{x}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}}\, s_{\bar{x}}$$

ومن الملاحظ أن فترة الثقة تعتمد على توزيع t لذلك يمكن إجراء التطبيق باستخدام برنامج SPSS.

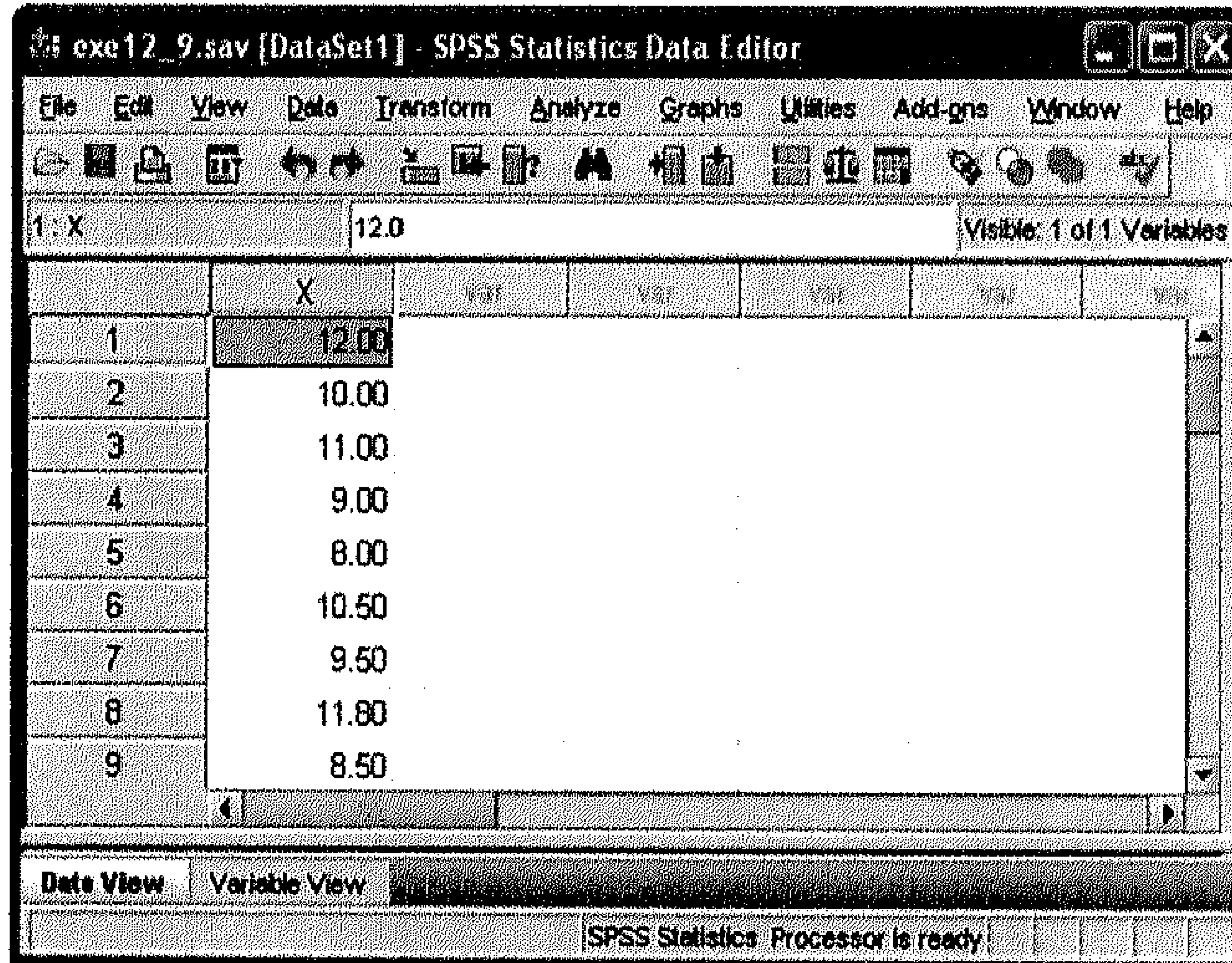
ولإجراء ذلك باستخدام برنامج SPSS سوف نتبع الخطوات التالية:

أولا نقوم بإدخال البيانات:

١- من نافذة Variable View نقوم بتعريف المتغير الكمي X



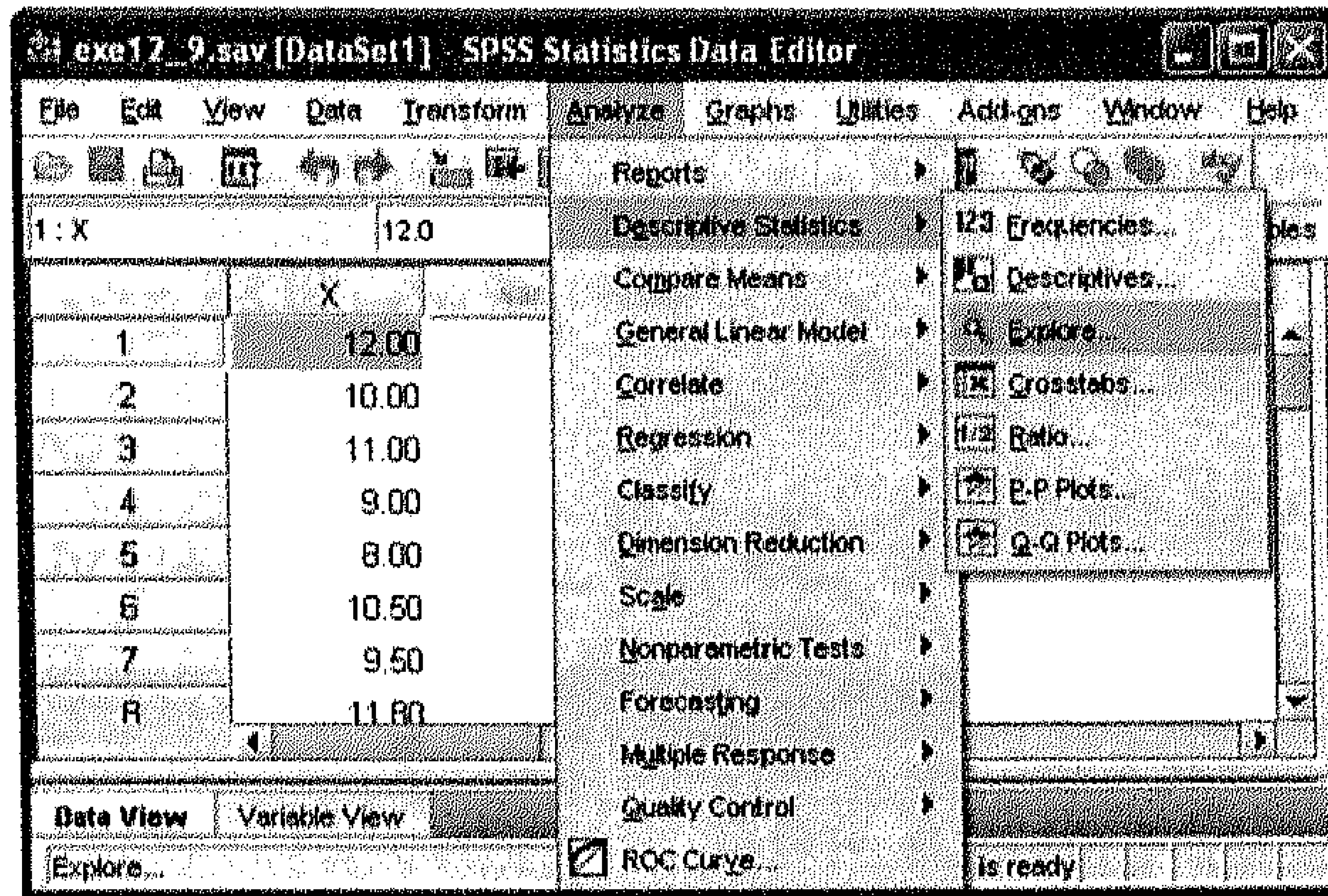
٢- ثم ننتقل لنافذة Data View ونقوم بإدخال البيانات الخاصة بالمتغير



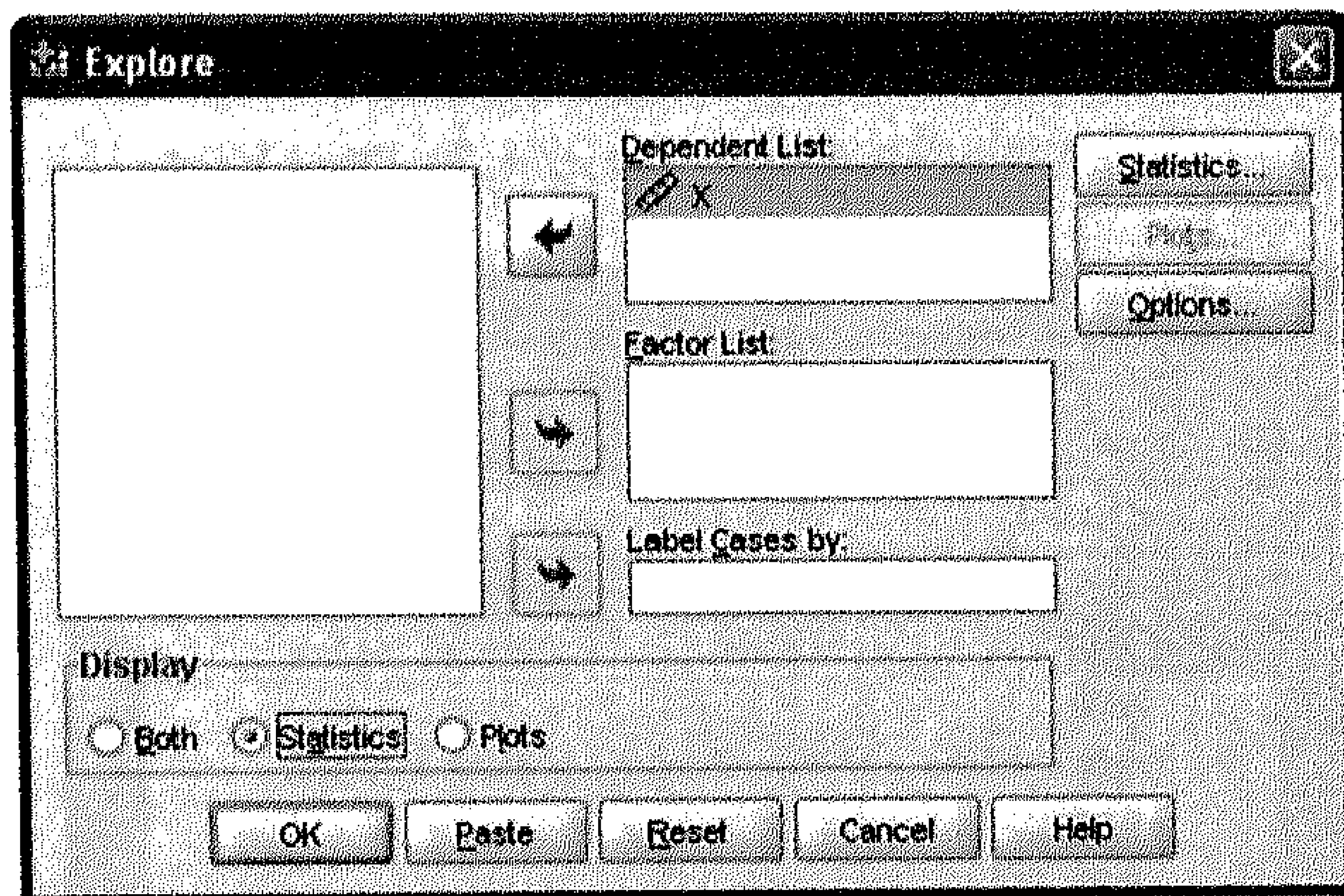
ثانياً تحديد فترة الثقة:

١- من قائمة Analyze نختار Descriptive Statistics

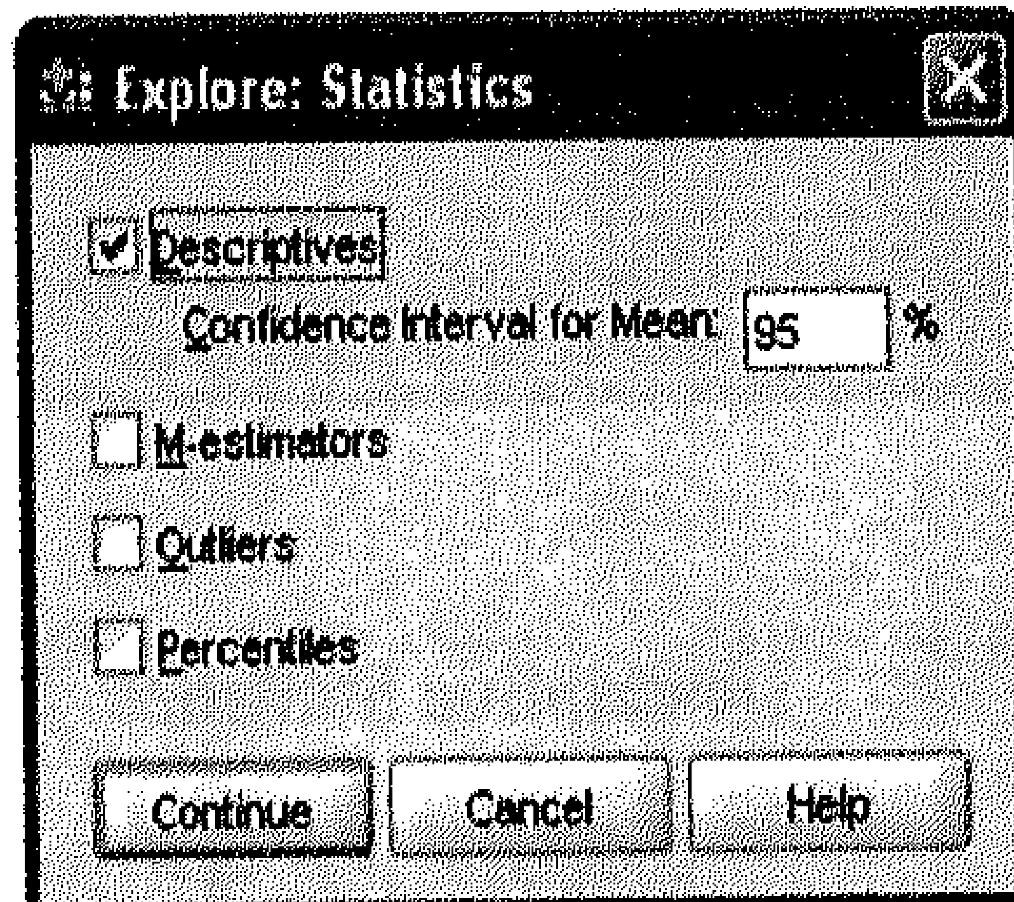
٢- تظهر قائمة منسدلة فنختار منها الأمر Explore



٣- تظهر نافذة جديدة بعنوان Explore فننقل المتغير X لقائمة Dependent List ونختار من قائمة Display الاختيار Statistics لعرض الإحصاءات فقط



٤- ثم نضغط على الاختيار Statistics تظهر الشاشة التالية



٥- فنجد أن الاختيار Descriptives محدد وأن Confidence Interval for Mean محدد أنها 95% فترة ثقة للمتوسط ويمكن تغييرها لأي قيمة نرغب فيها.

٦- ثم نختار Continue فنعود للشاشة السابقة نختار منها Ok فننتقل لنافذة المخرجات وتحتوي علي النتائج التالية

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
X	9	100.0%	0	.0%	9	100.0%

Descriptives

			Statistic	Std. Error
X	Mean		10.0333	.47111
	95% Confidence Interval for Mean	Lower Bound	8.9470	
		Upper Bound	11.1197	
	5% Trimmed Mean		10.0370	
	Median		10.0000	
	Variance		1.998	
	Std. Deviation		1.41333	
	Minimum		8.00	
	Maximum		12.00	
	Range		4.00	
	Interquartile Range		2.65	
	Skewness		.049	.717
	Kurtosis		-1.251	1.400

الجدول الأول يعطي عدد القيم للمتغير X وأيضا عدد القيم المتطرفة. وهى هنا صفر.

الجدول الثاني يعطى عديد من الإحصاءات وهي:

- الوسط الحسابي لقيم المتغير X وهى $\bar{x} = 10.00$ وهو المقدر بنقطة لمتوسط المجتمع.

- %95 فترة ثقة للمتوسط وهى $8.9475 \le \mu \le 11.0525$
- الوسيط للبيانات $median = 10$ ومن الملاحظ أن الوسط الحسابي أكبر من الوسيط لذا فإن المنحني غير متماثل بل ملتوٍ ناحية اليمين.
- التباين للبيانات $Variance = 1.875$
- الانحراف المعياري $Std.Deviation = 1.36931$
- أقل قيمة في البيانات $Minimum = 8$
- أكبر قيمة في البيانات $Maximum = 12$
- المدى للبيانات $Range = 4$
- المدى الربيعي $Interquartile\ Range = 2.50$
- معامل الالتواء $Skewness = 0.049$ وقيمة معامل الالتواء موجبة لذا فإن المنحني ملتوٍ ناحية اليمين كما سبق الإشارة إليه.
- معامل التفرطح $Kurtosis = -1.200$ وواضح أن منحني التوزيع مفلطح.

تطبيق(١١-٢)

في مثال(١١-١٩): من بيانات عام 1983م تبين أن متوسط أعمار العاملين بإحدى الهيئات الحكومية هو 38 سنة وفى عام 2004م تم اختيار عينة من العاملين بهذه الهيئة حجمها 25 شخصاً فكانت أعمارهم كما يلي:

28	34	43	45	45	33	45	26	50	46	39	41	39
44	39	40	41	31	43	35	40	44	46	33	37	

هل تدل هذه المشاهدات على أن متوسط أعمار العاملين في عام 2004م قد اختلف عن متوسط أعمارهم في عام 1983م وذلك عند مستوى معنوية 0.05؟ علماً بأن أعمار العاملين بتلك الهيئة يتبع التوزيع الطبيعي.

الحل

من المعطيات $n = 25, \mu_0 = 38$ فان

١- الفروض الإحصائية

$$H_0: \mu = 38$$

$$H_1: \mu \neq 38$$

٢- مستوى المعنوية $\alpha = 0.05$

٣- إحصاء الاختبار: بما أن المجتمع يتبع التوزيع الطبيعي وتباينه غير معلوم، وحجم العينة صغير فإن إحصاء الاختبار هو

$$T = \frac{\bar{X} - \mu_0}{s_{\bar{x}}}$$

ولإجراء هذا الاختبار باستخدام برنامج SPSS سوف نتبع الخطوات التالية:

أولاً نقوم بإدخال البيانات:

سوف نقوم بإدخال البيانات للبرنامج كالتالي:

١- من قائمة Variable View نقوم بتعريف المتغير الكمي X

٢- ثم ننتقل لقائمة Data View ونقوم بإدخال البيانات الخاصة بالمتغير

exe12_19.sav [DataSet1] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

1 : X 28.0 Visible: 1 of 1 Variables

	X
1	28.00
2	34.00
3	43.00
4	45.00
5	45.00
6	33.00
7	45.00
8	26.00
9	50.00
10	46.00

Data View Variable View

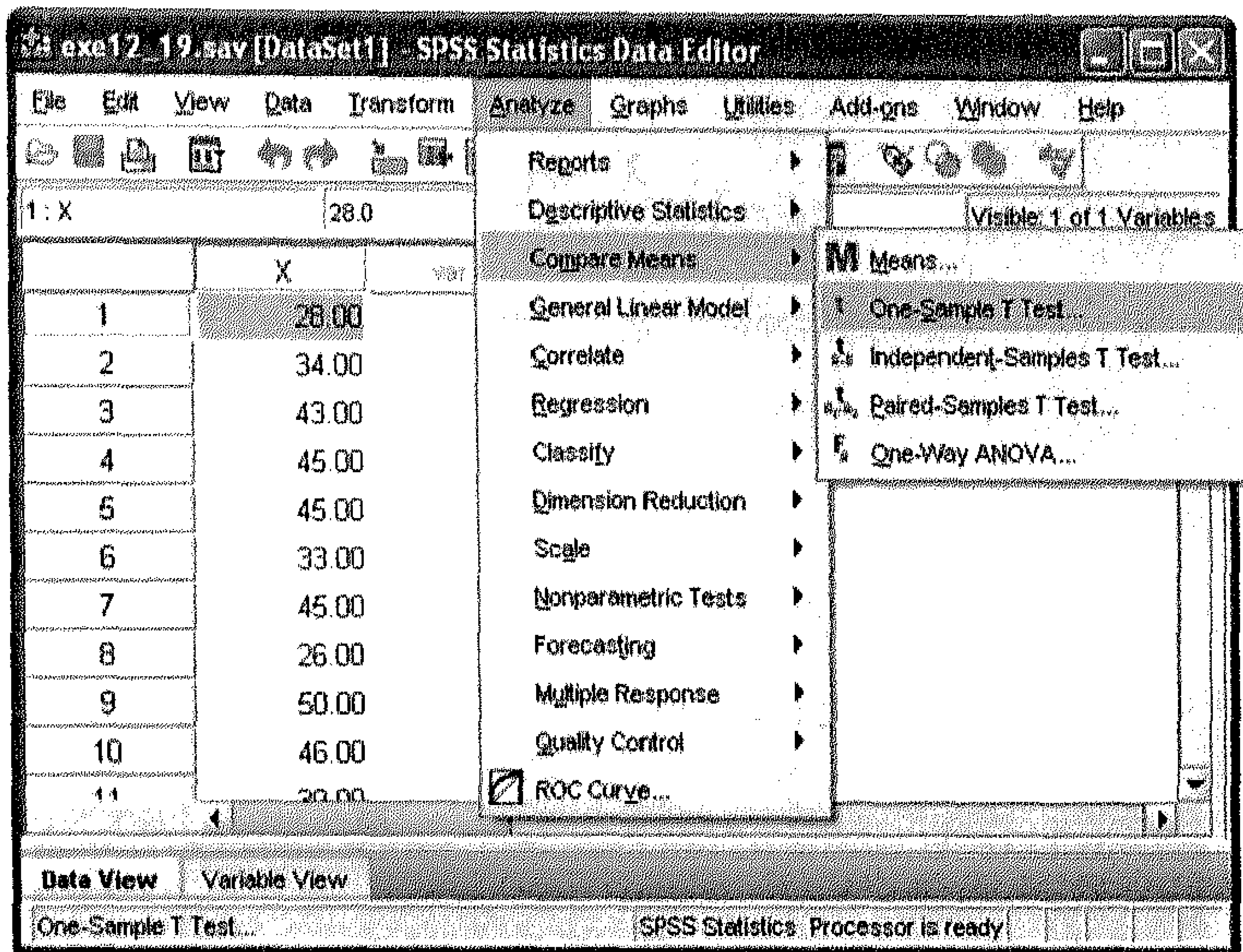
SPSS Statistics Processor is ready

ثانياً إجراء الاختبار:

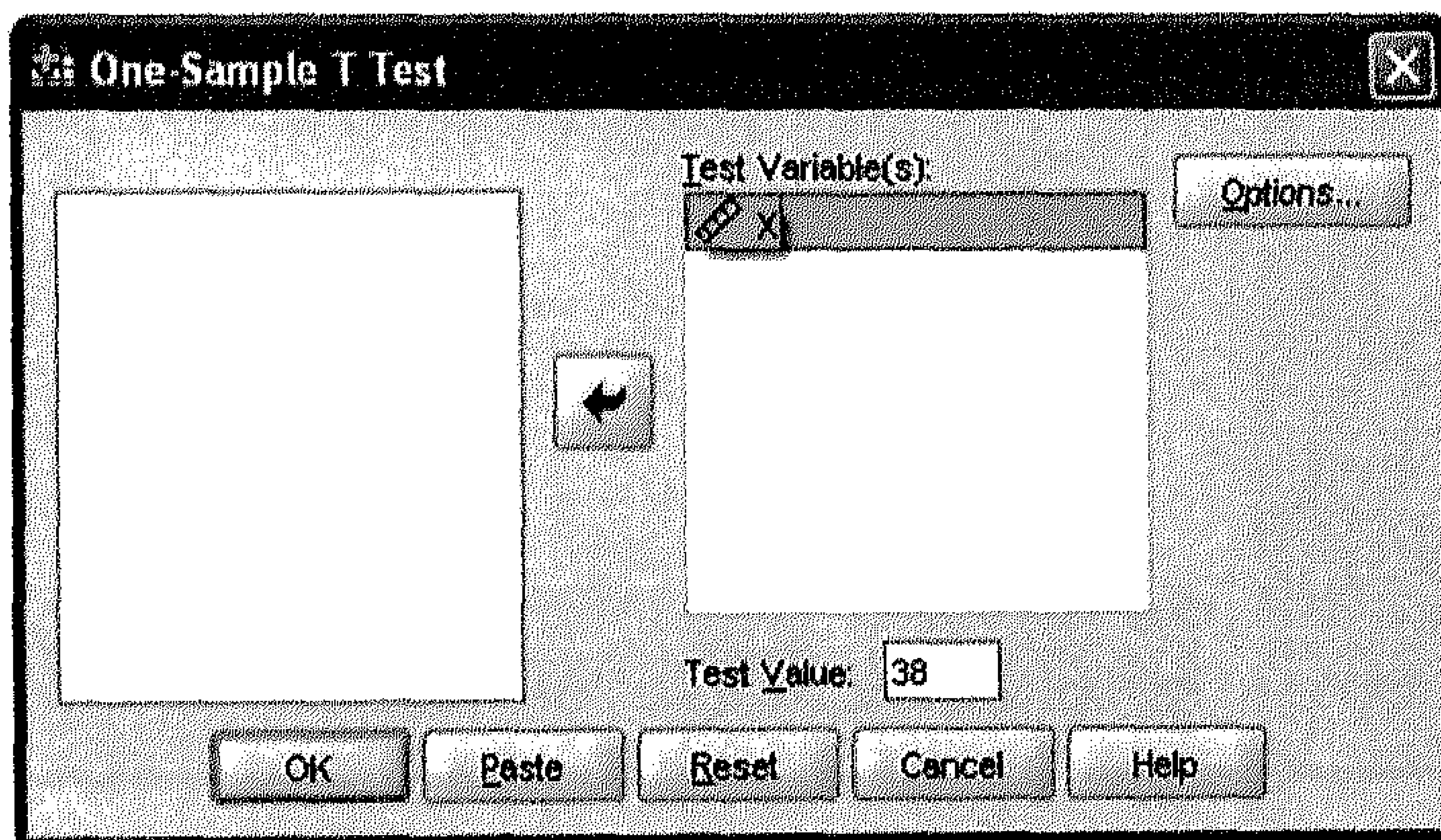
سوف نقوم بإجراء اختبار الفرض حول المتوسط باتباع الخطوات التالية:

١- من قائمة Analyze نختار Compare Means

٢- فتظهر قائمة منسدلة فنختار منها One Sample T Test

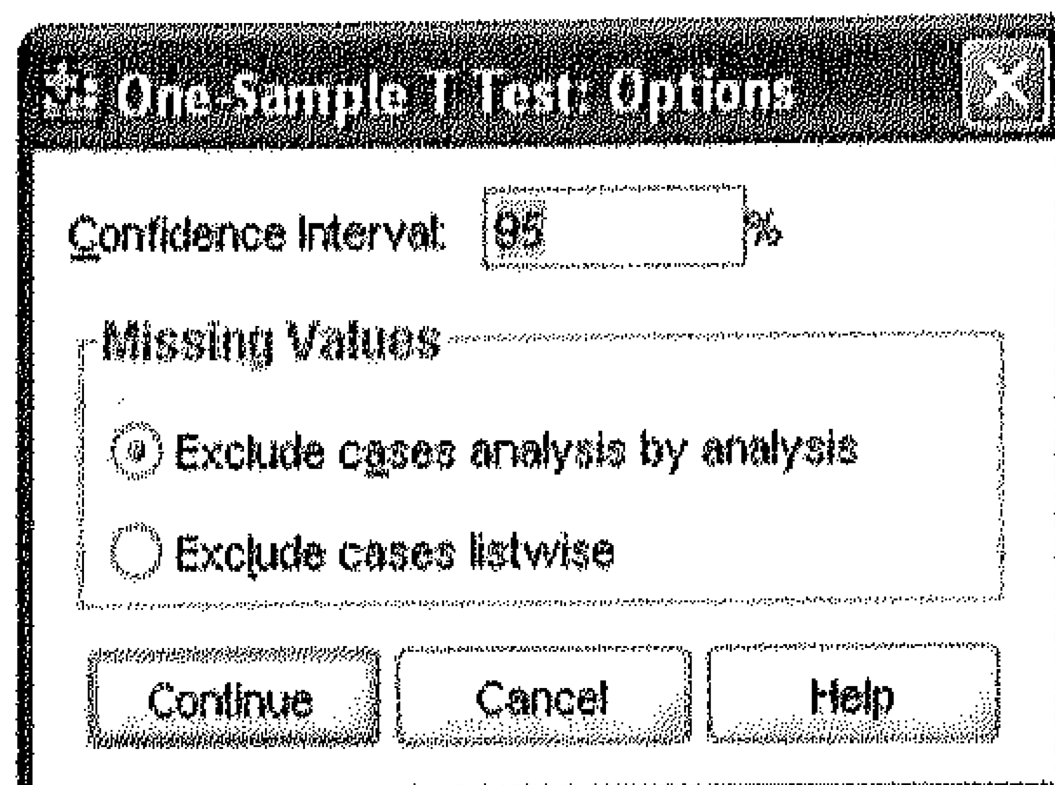


٣- ثم تظهر شاشة جديدة بعنوان One-Sample T Test فنقوم نقل المتغير X لقائمة Test variable(s)



٤- ونكتب قيمة $\mu_0 = 38$ فى خانة Test Value وهى القيمة المراد اختبار أن المتوسط يساوى تلك القيمة أم لا.

٥- نقوم بالضغط على الاختيار Options لتحديد مستوى المعنوية α



٦- ونلاحظ أن الاختيار $(1-)100\% = 95\%$ Confidence interval وهذا يعنى أن $\alpha = 0.05$ ويمكن تغييرها كما نريد وباختيار Continue نعود للشاشة السابقة فنختار Ok فننتقل لنافذة المخرجات Output View وتحتوي علي النتائج التالية

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
X	25	39.4800	6.06987	1.21397

One-Sample Test

	Test Value = 38					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
X	1.219	24	.235	1.48000	-1.0255	3.9855

ونجد أن الجدول الأول بعنوان One sample statistics يعطى الإحصاءات الخاصة بالعينة وهى:

- عدد عناصر العينة $n = 25$
- المتوسط الحسابي للعينة $\bar{x} = 39.48$
- الانحراف المعياري للعينة $s = 6.06987$
- الخطأ المعياري $S.E. = \frac{s}{\sqrt{n}} = 1.21397$

الجدول الثاني بعنوان one sample test ويعطى:

- إحصاء الاختبار $t = 1.219$
- درجة الحرية $\nu = n - 1 = 24$
- مساحة منطقة الرفض المحسوبة $sig = p - value = 2P[T \geq 1.219] = 0.235$
- الفرق بين قيمة متوسط العينة وقيمة μ_0 حيث $mean\ difference = \bar{x} - \mu_0 = 1.48$
- فترة 95% ثقة للفرق بين المتوسطين $\bar{x} - \mu$ والتي منها يمكن تعيين فترة الثقة للمتوسط.

ومن الجدول الثاني يمكن قبول أو رفض فرض العدم عن طريق:

١- قيمة إحصاء الاختبار $t = 1.219$ وحيث إن الاختبار ذو طرفين فإننا سوف نقوم بتعيين القيمة الحرجة $t_{\alpha/2} = t(n-1, \alpha/2) = t(24, 0.025) = 2.069$ ومنها نجد أن $-t_{\alpha/2} < t < t_{\alpha/2}$ وبالتالي فإن قيمةإحصاء الاختبار تقع في منطقة القبول فنقبل فرض العدم وهو أن متوسط أعمار العاملين في عام 2004م لا يختلف عن متوسط أعمارهم في عام 1983م.

٢- باستخدام القيمة الحرجة $sig = p - value = 0.235$ وواضح أن الاختبار ذو طرفين لذا سوف نقبل فرض العدم إذا كانت $Sig > \alpha$ ونرفض فرض العدم إذا كانت $Sig \leq \alpha$ ومن البيانات الموجودة بالجدول الثاني فنجد أن $Sig = 0.235 > \alpha = 0.05$ لذا سوف نقبل فرض العدم وهو أن متوسط أعمار العاملين في عام 2004م لا يختلف عن متوسط أعمارهم في عام 1983م.

أسئلة وتمارين (١١)

١- ما المقصود بالاستدلال الإحصائي؟

٢- ما المقصود بمستوى المعنوية؟

٣- عرف الفرض الإحصائي مع ذكر أنواعه.

٤- اذكر الخطوات الأساسية لإجراء الاختبار الإحصائي.

٥- ما هي العلاقة بين فترات الثقة واختبارات الفروض الإحصائية؟

٦- إذا تم قبول فرض العدم فهل هذا يعنى أن الاختبار معنوي؟

٧- ما المقصود بمنطقة الرفض ومنطقة القبول؟

٨- ما الشروط الواجب توافرها لإجراء الاختبارات التالية:

(أ) اختبار فرض حول توفر ظاهرة معينة في المجتمع.

(ب) اختبار فرض حول الوسط الحسابي للمجتمع.

(ج) اختبار فرض حول تباين المجتمع.

٩- حدد الفروض الإحصائية في العبارات التالية:

(أ) متوسط درجات طلاب كلية العلوم لا تختلف عن درجات طالبات كلية العلوم بجامعة الخرج.

(ب) نسبة الطلاب الحاصلين على معدل +A في مقرر 1080 إحص في كلية العلوم والدراسات الإنسانية بجامعة الخرج أقل من 10%

(ج) نسبة الوفيات في عام 1999م بسبب مرض السرطان قد انخفضت عنها في عام 1994م نتيجة للتقدم العلمي والطبي في طرق العلاج والاكتشاف المبكر للمرض.

(د) متوسط درجات طلاب كلية العلوم والدراسات الإنسانية في مقرر 1040 إحص أكبر من 72.

١٠- احسب حجم العينة العشوائية البسيطة اللازم اختيارها لتقدير المتوسط لظاهرة ما بخطأ لا يتعدى 0.9 وبدرجة ثقة 95% إذا علمت أن تباين تلك الظاهرة في المجتمع هو 2.5

١١- احسب حجم العينة العشوائية البسيطة اللازم سحبها من مجتمع مكون من 5000 شجرة من أشجار المانجو لتقدير متوسط إنتاج المانجو في ذلك المجتمع بخطأ لا يتعدى 5 وبدرجة ثقة 99% بفرض أن الانحراف المعياري لذلك المجتمع هو 45.

١٢- فى دراسة عن إنتاج أحد المصانع كان متوسط الانتاج اليومي خلال 80 يوماً هو 18.85 طناً. أوجد تقدير فترة الثقة للمتوسط الحقيقي μ لإنتاج المصنع اليومى عند درجة ثقة 95% علماً بأن الانحراف المعياري لإنتاج المصنع هو 5.55 طناً.

١٣- فى المثال السابق إذا علمت أن عدد الأيام التى تم تشغيل المصنع فيها هى $N = 500$ يوماً فأوجد تقدير فترة الثقة للمتوسط الحقيقي μ لإنتاج المصنع وذلك عند درجة ثقة 95%.

١٤- القراءات التالية تمثل الأوزان (بالجرام) لعبوات من الحلوى، يتم توزيعها بمعرفة أحد المصانع مجاناً للدعاية

64.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, 46.0

احسب القيمة العظمى للخطأ فى تقدير المتوسط للوزن، عند درجة ثقة 99% ثم أوجد 99% فترة الثقة للمتوسط الحقيقى μ لعبوات المصنع. مع العلم بأن أوزان العبوات يتبع التوزيع الطبيعي.

١٥- أراد باحث بأحد مصانع المصابيح الكهربائية أن يوجد فترة الثقة لمتوسط عمر المصباح الكهربائي الذى ينتجه المصنع فاختار لذلك عينة عشوائية مكونة من 9 مصابيح وقام بتسجيل أعمارها من لحظة إضاءتها حتى احترقت فوجد أن متوسط أعمارها هو 1000 ساعة، وكان الباحث على علم بأن أعمار هذه المصابيح تتبع التوزيع الطبيعي بانحراف معياري قدره 68 ساعة. فأوجد 99% فترة ثقة لمتوسط عمر المصابيح الكهربائية التى ينتجها هذا المصنع.

١٦- أخذت عينة مكونة من 81 حشرة من نوع معين بإحدى المناطق فوجد أن متوسط عمر الحشرة من هذا النوع والمحسوب من العينة هو 300 ساعة وكذلك وجد أن الانحراف المعياري المحسوب من العينة هو 27 ساعه فأوجد 95% فترة ثقة لمتوسط عمر الحشرة من هذا النوع.

١٧- لتقدير متوسط تركيز النحاس فى نبات وجد على ضفاف نهر النيل بمدينة المنصورة بمصر، أخذت عينة مكونة من 16 نبتة من هذا النبات وتم تحليلها لمعرفة تركيز النحاس فى هذا النبات فحصلنا على النتائج التالية: 20 ,19 ,60 ,25 ,66 ,35 ,43 ,38 ,50 ,8 ,14 ,27 ,18 ,34 ,3 ,4 فأوجد 99% فترة ثقة لمتوسط تركيز النحاس فى هذا النوع من النبات علماً بأن تركيز النحاس فى هذا النبات يتبع التوزيع الطبيعي.

١٨- أخذت عينة عشوائية مكونة من 9 أسر فى منطقة بها 5000 أسرة فكان عدد أفراد أسر العينة كما يلي:

7 ,3 ,6 ,5 ,4 ,8 ,5 ,3 ,4 فأوجد

(أ) 99% فترة ثقة لمتوسط عدد أفراد الأسرة بتلك المنطقة.

(ب) 99% فترة ثقة لعدد الأفراد الكلي بتلك المنطقة.

علماً بأن عدد أفراد الأسرة فى تلك المنطقة يتبع التوزيع الطبيعي.

١٩- احسب حجم العينة اللازم سحبها من مجتمع لتقدير نسبة المتعلمين فى ذلك المجتمع بخطأ لا يتعدى 0.02 وبدرجة ثقة 99% إذا كان معلوماً من دراسات سابقة أن تلك النسبة هى 0.6 تقريباً.

٢٠- أخذت عينة عشوائية مكونه من 900 فرداً من إحدى المدن فوجد أن منهم 500 مدخناً فأوجد 99% فترة ثقة لنسبة المدخنين فى هذه المدينة.

٢١- فى دراسة لاختبار صلاحية إحدى الطرق لعلاج مرض ما أخذت عينة مكونة من 400 شخصاً أبدى 136 شخصاً شعورهم بالراحة أثناء العلاج. احسب القيمة العظمى للخطأ فى التقدير d لنسبة التايید لطريقة العلاج عند درجة ثقة 95% وكذلك تقدير فترة الثقة للنسبة p عند نفس درجة الثقة.

٢٢- حل التمرين السابق إذا علم أن حجم المجتمع صغير ويتكون من 2000 شخصاً.

٢٣- لتحديد مقدار نسبة المصابين بضغط عال من بين الأشخاص البالغين فإن المطلوب هو تقدير حجم العينة حتى يمكننا التأكد باحتمال 0.95 من أن الخطأ لا يتجاوز 0.05 وذلك إذا علمنا أن قيمة $\hat{p}$ تساوى 0.2 فى الحالتين:

(أ) المجتمع ذو الحجم الكبير.

(ب) المجتمع ذو الحجم الصغير بحيث $N = 2000$

٢٤- أخذت عينة عشوائية حجمها 64 مصباحاً كهربائياً من إنتاج مصنع للمصابيح الكهربائية فوجد أن متوسط عمر المصباح الكهربائى فى العينة هو 570 ساعة بانحراف معيارى 60 ساعة. فأوجد 99% فترة ثقة لمتوسط عمر المصباح الكهربائي المنتج بواسطة هذا المصنع.

٢٥- أخذت عينة عشوائية مكونة من 8 حشرات من نوع معين فكانت أعمارها بالسنين هى 1.8, 1.9, 1.7, 2.1, 2.3, 1.6, 2.4, 2.2 والمطلوب

(أ) حساب 99% فترة ثقة لمتوسط عمر تلك الحشرة.

(ب) حساب 95% فترة ثقة لتباين عمر تلك الحشرة.

(ج) حساب 95% فترة ثقة للانحراف المعيارى لعمر الحشرة من ذلك النوع.

وذلك بفرض أن عمر الحشرة من هذا النوع يتبع التوزيع الطبيعى.

٢٦- أخذت عينة عشوائية مكونة من 600 من المدخنين فوجد أن 210 منهم مصابين بمرض معين. احسب 99% فترة ثقة لنسبة المدخنين المصابين بهذا المرض.

٢٧- أخذت عينة عشوائية مكونة من 400 طالب من طلاب جامعة سلمان بن عبدالعزيز، فكان عدد الناجحين فى مقرر معين هو 260 طالباً. احسب 95% فترة ثقة لنسبة الناجحين فى هذا المقرر بجامعة سلمان بن عبدالعزيز.

٢٨- سحبت عينة عشوائية بسيطه مكونة من 10 أفدنة من منطقة بها 6000 فداناً مزروعه قمحاً فكان إنتاج تلك الأفدنة بالكجم هو 890, 880, 870, 930, 920, 800, 800, 900, 960, 100 فاحسب

(أ) 99% فترة ثقة لمتوسط إنتاجية الفدان من القمح بتلك المنطقة.

(ب) 99% فترة ثقة لإجمالي الإنتاج من القمح بتلك المنطقة.

٢٩- احسب حجم العينة اللازم سحبها من مجتمع مكون من 20000 شخصاً لتقدير نسبة المدخنين بخطأ لا يتعدى 0.03 وبدرجة ثقة 0.99 بفرض أن نسبة المدخنين فى المجتمع تقع بين 85% ,65%.

٣٠- احسب حجم العينة اللازم سحبها من مجتمع لتقدير المتوسط لظاهرة معينة بخطأ لا يتعدى 0.2 بدرجة ثقة 0.95 بفرض أن تباين تلك الظاهرة فى المجتمع 1600.

٣١- أخذت العينة 1.3 ,2.1 ,1.7 ,2.3 ,0.07 ,1.1 ,1 من توزيع طبيعي $N(\mu, 1)$ اختبر الفرضيات التالية عند مستوى معنوية 0.05:

(أ) $H_0: \mu = 1.3$ مقابل $H_1: \mu \neq 1.3$

(ب) $H_0: \mu = 1.4$ مقابل $H_1: \mu < 1.4$

(ج) $H_0: \mu = 1.2$ مقابل $H_1: \mu > 1.2$

(د) $H_0: \sigma^2 = 1.242$ مقابل $H_1: \sigma^2 \neq 1.242$

(هـ) $H_0: \sigma = 1.5$ مقابل $H_1: \sigma \neq 1.5$

٣٢- أخذت عينة عشوائية تتكون من $n = 16$ سمكة وكان متوسط الوزن هو $\bar{X} = 158$ جرام وكان الانحراف المعياري للمجتمع معلوماً ويساوى 5 جرام. اختبر الفرض القائل بأن متوسط المجتمع الذي أخذت منه العينة يختلف عن 60 جرام وذلك عند مستوى معنوية $\alpha = 0.05$ علماً بأن الوزن يخضع للتوزيع الطبيعي.

٣٣- يدعى صاحب مزرعة أسماك أن متوسط وزن السمكة من إنتاج مزرعته لا يقل عن 1500 جم اختيرت عينة عشوائية مكونة من 9 سمكات من المزرعة فوجد أن متوسط الوزن لها هو $\bar{X} = 1400$ جم بانحراف

معياري s = 100 جم هل يمكن تأييد ادعاء صاحب المزرعة عند مستوى معنوية $\alpha = 0.01$ علماً بأن أوزان الأسماك تخضع للتوزيع الطبيعي.

٣٤- إذا كان معلوماً أن جسم الإنسان البالغ يحتاج يومياً في المتوسط إلى 800 مليجراماً من الكالسيوم لكي يقوم بوظائفه خير قيام. ويعتقد أحد علماء التغذية أن الأفراد ذوي الدخل المنخفض لا يستطيعون تحقيق هذا المتوسط. ولاختبار ذلك تم اختيار عينة من 50 شخصاً بالغاً من بين ذوي الدخل المنخفض فكان متوسط ما يتناولونه من كالسيوم يومياً هو 755.3 مليجراماً والانحراف المعياري هو 239.3 مليجراماً. فهل تدل هذه النتائج على أن متوسط ما يتناوله الأشخاص البالغون من ذوي الدخل المنخفض من كالسيوم يقل عن 800 مليجرام؟ استخدم مستوى معنوية 0.05.

٣٥- إذا كانت أعمار طلاب إحدي الجامعات تتبع التوزيع الطبيعي وادعى أحد الباحثين الاجتماعيين بأن متوسط عمر الطالب في هذه الجامعة هو 24 سنة، ولاختبار هذا الادعاء اختيرت عينة من 20 طالباً فكان متوسط أعمارهم هو 23 سنة والانحراف المعياري هو 3 سنوات. فهل معنى ذلك أن متوسط عمر طلاب الجامعة يختلف عن 24 سنة؟ استخدم مستوى معنوية 0.01.

٣٦- إذا كان متوسط أعمار إطارات السيارات المنتجة بواسطة أحد المصانع تتبع توزيعاً طبيعياً ويدعى صاحب المصنع أن متوسط أعمار هذه الإطارات هو 24 شهراً ولاختبار صحة هذا الادعاء اختيرت عينة عشوائية حجمها عشرة إطارات وقيست أعمارها بالشهور فكانت كما يلي: 27.6, 28.7, 34.7, 29.0, 22.9, 34.7, 36.5, 30.2, 29.4, 29.6 فهل تدل هذه البيانات على أن متوسط أعمار الإطارات اقل من 24 شهراً؟ استخدم مستوى معنوية 0.01

٣٧- إذا كانت أسعار إحدى السلع تتبع توزيعاً طبيعياً بانحراف معياري قدره 4 ريالات سعودية وكان متوسط سعر السلعة عام 1987م هو 38.5 ريالاً للكيلوجرام، وفى عام 1995م تم اختيار عينة من أسعار هذه السلعة حجمها 15 فكان وسطها الحسابي هو 40 ريالاً. هل تدل هذه البيانات على اختلاف متوسط سعر السلعة في العامين؟ استخدم مستوى معنوية 0.01

٣٨- إذا كانت نسبة المدخنين في إحدي المدن عام 1974م هي 27.8% وفى عام 1999م اختيرت عينة من سكان هذه المدينة حجمها 1200 شخصاً فكان من بينهم 300 شخصاً من المدخنين. فهل تدل هذه النتائج على انخفاض نسبة المدخنين بين عامي 1974، 1999م؟ استخدم مستوى معنوية 0.05.

٣٩- من مجتمع يحتوى 10000 شخصاً أخذت عينة مكونة من 450 فرداً فوجد أن 100 شخصاً منهم من غير السعوديين. فهل تدل هذه البيانات على أن نسبة غير السعوديين بهذا المجتمع تزيد عن %30 من السكان؟ استخدم مستوى معنوية 0.05

٤٠- لاختبار الفرض بأن تباين المجتمع $120 = \sigma^2$ اختيرت عينة عشوائية حجمها 25 مفردة فوجد أن تباينها هو 127.07. استخدم مستوى معنوية 0.01 لاختبار هذا الفرض علماً بأن المجتمع يتبع التوزيع الطبيعي.

٤١- ادعى أحد أعضاء هيئة التدريس بقسم الإعداد العام بكلية العلوم والدراسات الإنسانية بجامعة سلمان بن عبدالعزيز بأن درجات طلاب الكلية في مقرر 103 سلم يزيد عن 74 درجة فتم اختيار عينة من بين الطلاب حجمها 50 طالباً فكان متوسط درجاتهم 79 درجة بانحراف معياري 4 درجات. فهل يمكن القول بأن هذا الادعاء صحيح؟ استخدم مستوي معنوية 0.05

٤٢- ادعى أحد أعضاء وحدة الإنتاج بمصنع للأجهزة الكهربائية بأن نسبة الأجهزة التالفة المنتجة يومياً هي 0.001 فتم اختيار عينة من الإنتاج اليومي للمصنع حجمها 400 جهازاً فكان عدد الأجهزة التالفة في العينة هو 3 أجهزة. فهل يمكن القول بأن هذا الادعاء صحيح؟ استخدم مستوي معنوي 0.01

٤٣- من مجتمع يحتوى 1000 شخص أخذت عينة مكونة من 250 فرداً فوجد أن 50 شخصاً منهم من ذوي المؤهلات العليا. فهل تدل هذه البيانات على أن نسبة ذوي المؤهلات العليا في هذا المجتمع تزيد عن %25 من السكان؟ استخدم مستوى معنوية 0.01

٤٤- إذا كان الدخل الشهري للأفراد السعوديين يتبع توزيعاً طبيعياً بانحراف معياري قدره 2000 ريال سعودي وكان متوسط الدخل في عام 1427هـ هو 6000 ريالاً، وفى عام 1430هـ تم اختيار عينة من الأفراد السعوديين حجمها 15 فرداً وتم الاستعلام عن دخلهم الشهري وكانت البيانات كالتالي:

8000	7600	4500	8300	12300	8900	11500
8950	7570	16500	9450	5700	10400	6450
9900						

هل تدل هذه البيانات على ارتفاع مستوي الدخل في عام 1430هـ عن عام 1427 هـ؟ استخدم مستوى معنوية 0.01

جدول الأرقام العشوائية

TABLE I: RANDOM NUMBERS

63271	59986	71744	51102	15141	80714	58683	93108	13554	79945
88547	09896	95436	79115	08303	01041	20030	63754	08459	28364
55957	57243	83865	09911	19761	66535	40102	26646	60147	15704
46276	87453	44790	67122	45573	84358	21625	16999	13385	22782
55363	07449	34835	15290	76616	67191	12777	21861	68689	03263
69393	92785	49902	58447	42048	30378	87618	26933	40640	16281
13186	29431	88190	04588	38733	81290	89541	70290	40113	08243
17726	28652	56836	78351	47327	18518	92222	55201	27340	10493
36520	64465	05550	30157	82242	29520	69753	72602	23756	54935
81628	36100	39254	56835	37636	02421	98063	89641	64953	99337
84649	48968	75215	75498	49539	74240	03466	49292	36401	45525
63291	11618	12613	75055	43915	26488	41116	64531	56827	30825
70502	53225	03655	05915	37140	57051	48393	91322	25653	06543
06426	24771	59935	49801	11082	66762	94477	02494	88215	27191
20711	55609	29430	70165	45406	78484	31639	52009	18873	96927
41990	70538	77191	25860	55204	73417	83920	69468	74972	38712
72452	36618	76298	26678	89334	33938	95567	29380	75906	91807
37042	40318	57099	10528	09925	89773	41335	96244	29002	46453
53766	52875	15987	46962	67342	77592	57651	95508	80033	69828
90585	58955	53122	16025	84299	53310	67380	84249	25348	04332
32001	96293	37203	64516	51530	37069	40261	61374	05815	06714
62606	64324	46354	72157	67248	20135	49804	09226	64419	29457
10078	28073	85389	50324	14500	15562	64165	06125	71353	77669
91561	46145	24177	15294	10061	98124	75732	00815	83452	97355
13091	98112	53959	79607	52244	63303	10413	63839	74762	50289
73864	83014	72457	22682	03033	61714	88173	90835	00634	85169
66668	25467	48894	51043	02365	91726	09365	63167	95264	45643
84745	41042	29493	01836	09044	51926	43630	63470	76508	14194
48068	26805	94595	47907	13357	38412	33318	26098	82782	42851
54310	96175	97594	88616	42035	38093	36745	56702	40644	83514
14877	33095	10924	58013	61439	21882	42059	24177	58739	60170
78295	23179	02771	43464	59061	71411	05697	67194	30495	21157
67524	02865	39593	54278	04237	92441	26602	63835	38032	94770
58268	57219	68124	73455	83236	08710	04284	55005	84171	42596
97158	28672	50685	01181	24262	19427	52106	34308	73685	74246
04230	16831	69085	30802	65559	09205	71829	06489	85650	38707
94879	56606	30401	02602	57658	70091	54986	41394	60437	03195
71446	15232	66715	26385	91518	70566	02888	79941	39684	54315
32886	05644	79316	09819	00813	88407	17461	73925	53037	91904
62048	33711	25290	21526	02223	75947	66466	06232	10913	75336
84534	42351	21628	53669	81352	95152	08107	98814	72743	12849
84707	15885	84710	35866	06446	86311	32648	88141	73902	69981
19409	40868	64220	80861	13860	68493	52908	26374	63297	45052

جدول التوزيع الطبيعي القياسي

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

Table II Cumulative Standard Normal Distribution: $P[Z \le z]$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5039	0.5079	0.5119	0.5159	0.5199	0.5329	0.5279	0.5318	0.5358
0.1	0.5398	0.5438	0.5478	0.5517	0.5558	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5792	0.5832	0.5871	0.5909	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6949	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7674	0.7704	0.7734	0.7764	0.7794	0.7824	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8079	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8339	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8769	0.8789	0.8810	0.8829
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8979	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9369	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9829	0.9834	0.9838	0.9842	0.9846	0.9849	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9889
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9939	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9959	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9983	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9989	0.9989
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

جدول توزيع مربع كاي

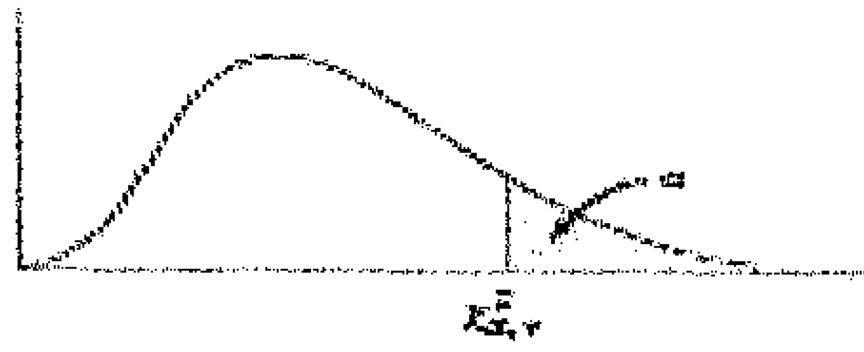


Table III Percentage Points $\chi^2_{\alpha,\nu}$ of the Chi-Squared Distribution

ν \ α	0.995	0.990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.65
28	12.46	13.57	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
19	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

جدول توزيع t

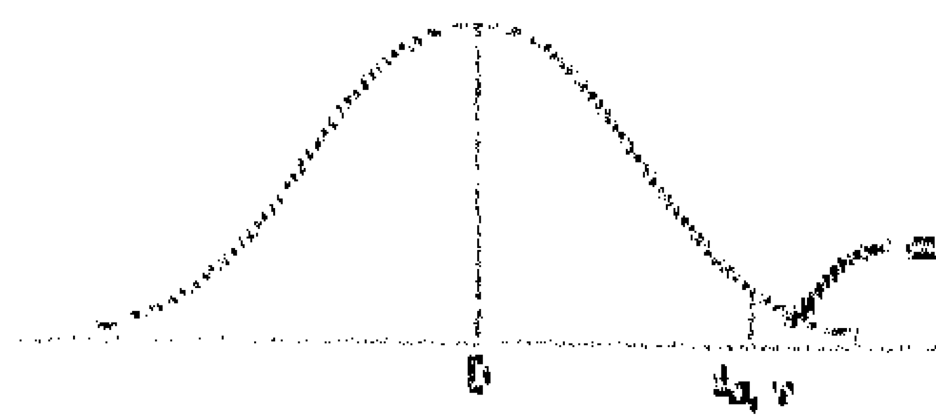


Table IV Percentage Points $t_{\alpha,\nu}$ of the t-Distribution

ν \ α	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.82	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

المراجع

REFERENCES

أولا: المراجع العربية

- أبو صالح، محمد صبحي وعدنان محمد عوض. مقدمة في الإحصاء. دار جون وايلي وأبنائه- نيويورك. ١٩٨٣م.
- أبو صالح، عدنان محمد عوض. مقدمة في الإحصاء/ مبادئ وتحليل باستخدام SPSS . دار المسيرة للنشر والتوزيع والطباعة – عمان – الاردن. ٢٠١٠م.
- أبو عمه، عبد الرحمن والحسيني راضي، محمود هندي. مقدمة في المعاينة الإحصائية. دار المريخ للنشر – الرياض – المملكة العربية السعودية. ١٤١٥هـ.
- الجمعة، علي محمد. مقدمة في التحليل الإحصائي. مكتبة الشقري – الرياض – المملكة العربية السعودية. ١٤٢٥هـ.
- الشافعي، فتحي شريف. الدليل العلمي للتحليل الإحصائي باستخدام برنامج SPSS 14.0 – دار الكتب العلمية للنشر والتوزيع – مصر. ٢٠٠٦م.
- العباسي، عبدالحميد. التحليل الإحصائي باستخدام SPSS، معهد الدراسات والبحوث الاحصائية – جامعة القاهرة- مصر. ١٩٩٩م.

- الشيحة، عبد الله وعدنان بري. مقدمة في الإحصاء والاحتمالات وتطبيقاتها باستخدام إكسل. مكتبة الشقري – الرياض – المملكة العربية السعودية. ١٤٢٩ هـ.
- الصياد، جلال مصطفى. نظرية الاحتمالات. الطبعة الخامسة. دار حافظ – دمشق – سورية. ١٤٢٣هـ.
- باشيوه، لحسن. الإحصاء وتطبيقاته على الحزمة الإحصائية SPSS. دار الوراق للخدمات الحديثة. ٢٠١١م
- بري، عدنان ماجد وآخرون. مبادئ الإحصاء والاحتمالات. الطبعة الثالثة. النشر العلمي والمطابع، جامعة الملك سعود– الرياض – المملكة العربية السعودية. ١٤١٨هـ.
- بري، عدنان ماجد ومحمود هندي. مبادئ الإحصاء والاحتمالات مع حل الأمثلة باستخدام ميكروسوفت إكسل. مكتبة الشقري– الرياض – المملكة العربية السعودية. ١٤٢٦هـ.
- عاشور، سمير كامل– سالم، سامية أبو الفتوح. العرض والتحليل الإحصائي باستخدام SPSSWIN، الجزء الأول: المدخل والأساسيات، معهد الدراسات والبحوث الإحصائية– جامعة القاهرة – مصر. ٢٠٠٨م
- عاشور، سمير كامل– سالم، سامية أبو الفتوح. العرض والتحليل الإحصائي باستخدام SPSS/PC+، الجزء الثاني: الإحصاء التطبيقي المتقدم، معهد الدراسات والبحوث الإحصائية– جامعة القاهرة – مصر. ٢٠٠٩م
- عاشور، سمير كامل– سالم، سامية أبو الفتوح. العرض والتحليل الإحصائي باستخدام SPSSWIN، الجزء الثالث: تصميم التجارب وتحليلها، معهد الدراسات والبحوث الإحصائية– جامعة القاهرة – مصر. ٢٠٠٩م
- عبدالفتاح، عز حسن سيد. مقدمة في الإحصاء الوصفي والإحصاء الاستدلالي باستخدام SPSS. خوارزم العلمية– جده – المملكة العربية السعودية. ٢٠٠٨م.
- عاروري، فتحي. المعاينة الإحصائية طرقها واستخداماتها. الأكاديميون للنشر والتوزيع – عمان – الاردن. ٢٠١٠م.

- عودة، أحمد عودة عبدالمجيد ومنصور عبدالرحمن القاضي. الإحصاء الوصفي والاستدلالي. مكتبة الفلاح للنشر والتوزيع. الكويت. ١٤٢٢هـ.
- كنجو، أنيس إسماعيل. الإحصاء والاحتمال. مكتبة العبيكان- الرياض - المملكة العربية السعودية. ١٤٢١هـ
- كنجو، أنيس إسماعيل وعبد الرحمن أبو عمة . مدخل إلى نظرية الاحتمال وتطبيقاتها. جامعة الملك سعود- الرياض - المملكة العربية السعودية. ١٤٢١هـ.
- محمود، قطب عبدالحميد. مقدمة في الإحصاء والاحتمالات. خوارزم العلمية. جده- المملكة العربية السعودية. ٢٠٠٦م.
- مشهور، أحمد فوزي. الإحصاء والاحتمالات. مذكرة محاضرات. كلية العلوم - جامعة المنصورة- مصر.
- هندي، محمود وخلف سلطان. مفاهيم لطرق التحليل الإحصائي. مكتبة الرشد- الرياض - المملكة العربية السعودية. ١٤٢٥هـ.

ثانيا المراجع الانجليزية

Chatfield, C. and Zidek, J. Mathematical Statistics. Chapman & Hall/Krc, 2000.

DeGroot, M.H and Schervish, M. J. Probability and Statistics, 4th Edition. Pearson Education, 2012.

Dekking, F.M., Kraaikamp, C. , Lopuhaä, H.P. and Meester, L.E. A Modern Introduction to Probability and Statistics: Understanding Why and How. Springer-Verlag London Limited, 2005.

Devore, J.L. Probability and Statistics for Engineering and the Sciences: Enhanced. 7th Edition. Brooks/Cole, Cengage Learning, USA, 2009,

Gupta, V. SPSS for Beginners. VJBooks Inc., 1999.

http://www.amazon.co.uk/Crash-Course.../dp/1405145315

Ho, R. 2006. Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. Chapman & Hall/CRC, Taylor & Francis Group, New York.

Landau, S. and Everitt, B.S. A Handbook of Statistical Analyses using SPSS. A CRC Press Company , Washington, D.C., 2004.

Levesque, R. SPSS Programming and Data Management, 2nd Edition, A Guide for SPSS and SAS Users, 2005.

Lomax, R.G. An introduction to statistical concepts for education and behavioral sciences. Lawrence Erlbaum Associates, Inc., 2001.

Mahajan, S. Street-Fighting Mathematics: The Art of Educated Guessing and Opportunistic Problem Solving. The MIT, Cambridge, Massachusetts, London, England, 2010.

Mark L.B., David M.L., Timothy C.K., and David F.S. Basic Business Statistics - Concepts and Applications, 12nd Edition. Pearson Education, Inc., publishing as Prentice Hall, 2012.

Hodges, J.L. and Lehmann, E.L. Basic Concepts of Probability and Statistics, 2nd edition (Classics in Applied Mathematics). The Society for Industrial and Applied Mathematics, 2005.

Norusis, M. SPSS 13.0 Guide to Data Analysis. Techniques of Population Analysis. By G. W. Barclay, Wiley and Sons.

Rosenkrantz, W.A. Introduction to Probability and Statistics for Science, Engineering, and Finance. CRC Press, Taylor & Francis Group, New York, 2009.

Ross, S.M. Introduction to Probability Models. 9th Edition. Academic Press, San Diego, 2007.

Ross, S.M. Introduction to Probability and Statistics for Engineers and Scientists. (4th Edition) Elsevier, 2009.

Ross, S.M. A First Course in Probability. Prentice- Hall, Inc, 1998.

Scheaffer, R.L. and Young, L.J. Introduction to Probability and Its Applications, 3rd Edition. Brooks/Cole, Cengage Learning, United States, 2010.

Soong, T. T. Fundamental of Probability and Statistics for Engineers. John Wiley & Sons, Ltd., 2004.

Stirzaker, D. Elementary Probability, 2nd Edition. Cambridge University Press, 2003.

Suhov, Y. and Kelbert, M. Probability and Statistics by Example, Volume I. Basic Probability and Statistics. Cambridge University Press, 2005.

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. Probability & Statistics for Engineers & Scientists. 8th Edition. Pearson Prentice Hall, Pearson Education, Inc., 2007.

Weiss, N.A. Elementary Statistics. 8th Edition. Pearson Education, Boston, 2012.

ثبت المصطلحات

أولاً: عربي- إنجليزي

أ

اتحاد حادثتين	Union of two events
اختبارات الفروض	Hypotheses tests
اختبارات الفروض الإحصائية	Statistical testing of Hypotheses
افتراضي	Hypothetical
الاحتمال الشرطي	Conditional probability
الاحتمالات	Probability
الارتباط	Correlation
الاستدلال الإحصائي	Statistical inference
الالتواء	Skewness
الانحدار الخطي البسيط	Simple linear regression
الانحراف المعياري	Standard deviation
الإحصاء الاستدلالي	Inferential Statistics
الإحصاء الوصفي	Descriptive Statistics

Bar Chart	الأعمدة البيانية
Grouped data	البيانات المبوبة
Row data	البيانات المفردة
Permutation	التباديل
Variance	التباين
Inductive reasoning	التبرير الاستقرائي
Deductive reasoning	التبرير الاستنباطي
Relative dispersion	التشتت النسبي
Mathematical definition of probability	التعريف الرياضي للاحتمال
Classical definition of probability	التعريف الكلاسيكي للاحتمال
Kurtosis	التفرطح
Estimation	التقدير
Combination	التوافيق
Probability distribution	التوزيع الاحتمالي
Sampling distribution of the sample variance	التوزيع الاحتمالي لتباين العينة
Sampling distribution of the sample proportion	التوزيع الاحتمالي للنسبة في العينة
Exponential distribution	التوزيع الأسي
Normal distribution	التوزيع الطبيعي
Standard Normal distribution	التوزيع الطبيعي القياسي
Uniform distribution	التوزيع المنتظم
Geometric distribution	التوزيع الهندسي
Hyper Geometric distribution	التوزيع الهندسي الزائدي
Mathematical expectation	التوقع الرياضي

الجداول التكرارية المزدوجة	Frequency cross tables
الحوادث المتنافية	Mutually exclusive events
الحوادث المستقلة	Independent events
الحادثة المكملة	Complements events
الدائرة البيانية	Pie Chart
السحب بإرجاع	Sampling with replacement
السحب بدون إرجاع	Sampling without replacement
العلاقة بين الحوادث	Relation between events
العينة الطبقية	Stratified random sample
العينة العشوائية البسيطة	Simple random sample
العينة العنقودية	Cluster sample
العينة المعيارية	Standard sample
العينة المنتظمة	Systematic sample
العينات المرتبة	Ordered sample
العينات غير المرتبة	Non – ordered samples
الفرق بين حادثتين	Difference between two events
الوحدات القياسية	Standard unit
إحصاء	Statistics

ب

بعض نظريات الاحتمالات	Some probability theorems
بيانات	Data
بيانات كمية	Quantitative data
بيانات وصفية	Qualitative data

ت

Random experiment	تجربة عشوائية
Dispersion	تشتت
Definition of probability	تعريف الاحتمال
Variability	تغير
Intersection of two events	تقاطع حادثتين
Confidence interval estimation for the variance	تقدير فترة الثقة للتباين
Confidence interval estimation for the mean	تقدير فترة الثقة للمتوسط
Confidence interval estimation for the proportion	تقدير فترة الثقة للنسبة
Equivalence of two events	تكافؤ حادثتين
Frequency	تكرار
Bernoulli distribution	توزيع برنوللي
Poisson distribution	توزيع بواسون
Student`s t distribution	توزيع ت
Binomial distribution	توزيع ذي الحدين
Negative binomial distribution	توزيع ذي الحدين السالب
Chi – square distribution	توزيع مربع كاي
Sampling distribution of the mean	توزيع المعاينة للوسط
Sampling distribution	توزيعات المعاينة
Cumulative frequency distribution	توزيعات تكرارية متجمعة

ث

confidence	ثقة

ح

Event	حادثة

Simple event	حادثة بسيطة
Sub event	حادثة جزئية
Random event	حادثة عشوائية
Sure event	حادثة مؤكده
Compound event	حادثة مركبة
Impossible event	حادثة مستحيلة
Determined	حتمية
Size of rejection region	حجم منطقة الرفض
Packages	حزم
Statistical packages	حزم إحصائية
Real	حقيقي

خ

Type I error	خطأ من النوع الأول
Type II error	خطأ من النوع الثاني

د

Distribution function	دالة التوزيع
Density function	دالة الكثافة

ذ

One tailed	ذو طرف واحد
Two tailed	ذو طرفين

ر

Quartiles	ربيعات

ش

شكل الانتشار	Scatter diagram

ط

طرق العد	Counting methods
طريقة المربعات الصغرى	Least square method

ع

عشيرات	Decimals
عينة	Sample

غ

غير محدود	finite

ف

فئات	Classes
فترة الثقة	Confidence interval
فراغ العينة	Sample space
فرض بديل	Alternative hypothesis
فرض العدم	Null hypothesis

ق

قاعدة ضرب الاحتمالات	The multiplication rule
قانون الاحتمال الكلي	Law of total probability
قيمة مركزية	Central value

م

مئينات	Percentages
مبادئ الاستدلال الإحصائي	Elementary statistical inference
متباينة	Inequality

Chebychev's inequality	متباينة تشيبيشف
Variable	متغير
Random variable	متغير عشوائي
Continuous random variable	متغير عشوائي متصل
Discrete random variable	متغير عشوائي متقطع
Quantitative variable	متغير كمي
Qualitative variable	متغير وصفي
Symmetric	متماثل
Weighted mean	متوسط موزون
Averages	متوسطات
Population	مجتمع
Finite	محدود
Frequency Histogram	مدرج تكراري
Relative frequency histogram	مدرج تكراري نسبي
Range	مدى
Level of significance	مستوي المعنوية
Observations	مشاهدات
Frequency polygon	مضلع تكراري
Parameters	معالم
Population parameters	معالم المجتمع
Rank coefficient of correlation for Spearman	معامل ارتباط الرتب لسبيرمان
Coefficient of variation	معامل الاختلاف
Coefficient of contingency	معامل الاقتران
Linear correlation coefficient of Pearson	معامل الانحدار الخطي لبيرسون

Coefficient of association	معامل التوافق
Measure of dispersion	مقاييس التشتت
Measures of central f tendency	مقاييس النزعة المركزية
Negative skewed	ملتوياً ناحية اليسار
Positively skewed	ملتوياً ناحية اليمين
Curve	منحنى
Rejection region	منطقة الرفض
Acceptance region	منطقة القبول
Mode	منوال

ن

Relative	نسبي
Semi interquartile range	نصف المدى الربيعي
Theorem	نظرية
Bayessian theorem	نظرية بيز
Central limit theorem	نظرية النهاية المركزية

و

Harmonic mean	وسط توافقي
Arithmetic mean	وسط حسابي
Geometric mean	وسط هندسي
Median	وسيط
Random variable description	وصف المتغير العشوائي

ثانياً: إنجليزي – عربي

A

Acceptance region	منطقة القبول
Alternative hypothesis	فرض بديل
Arithmetic mean	وسط حسابي
Averages	متوسطات

B

Bar Chart	الأعمدة البيانية
Bayessian theorem	نظرية بيز
Bernoulli distribution	توزيع برنوللي
Binomial distribution	توزيع ذي الحدين

C

Central limit theorem	نظرية النهاية المركزية
Central value	قيمة مركزية
Chebychev's inequality	متباينة تشيبيشف
Chi – square distribution	توزيع مربع كاي
Classes	فئات
Classical definition of probability	التعريف الكلاسيكي للاحتمال
Cluster sample	العينة العنقودية
Coefficient of association	معامل التوافق
Coefficient of contingency	معامل الاقتران
Coefficient of variation	معامل الاختلاف
Combination	التوافيق

Complements events	الحادثة المكملة
Compound event	حادثة مركبة
Conditional probability	الاحتمال الشرطي
Confidence	ثقة
Confidence interval	فترة الثقة
Confidence interval estimation for the mean	تقدير فترة الثقة للمتوسط
Confidence interval estimation for the proportion	تقدير فترة الثقة للنسبة
Confidence interval estimation for the variance	تقدير فترة الثقة للتباين
Continuous random variable	متغير عشوائي متصل
Correlation	الارتباط
Counting methods	طرق العد
Cumulative frequency distribution	توزيعات تكرارية متجمعة
Curve	منحني

D

Data	بيانات
Decimals	عشيرات
Deductive reasoning	التبرير الاستنباطي
Definition of probability	تعريف الاحتمال
Density function	دالة الكثافة
Descriptive Statistics	الإحصاء الوصفي
Determined	حتمية
Difference between two events	الفرق بين حادثتين
Discrete random variable	متغير عشوائي متقطع
Dispersion	تشتت

Distribution function	دالة التوزيع

E

Elementary statistical inference	مبادئ الاستدلال الإحصائي
Equivalence of two events	تكافؤ حادثتين
Estimation	التقدير
Event	حادثة
Exponential distribution	التوزيع الأسي

F

Finite	غير محدود
Finite	محدود
Frequency	تكرار
Frequency cross tables	الجداول التكرارية المزدوجة
Frequency Histogram	مدرج تكراري
Frequency polygon	مضلع تكراري

G

Geometric distribution	التوزيع الهندسي
Geometric mean	وسط هندسي
Grouped data	البيانات المبوبة

H

Harmonic mean	وسط توافقي
Hyper Geometric distribution	التوزيع الهندسي الزائدي
Hypotheses tests	اختبارات الفروض
Hypothetical	افتراضي

I

Impossible event	حادثة مستحيلة
Independent events	الحوادث المستقلة
Inductive reasoning	التبرير الاستقرائي
Inequality	متباينة
Inferential Statistics	الإحصاء الاستدلالي
Intersection of two events	تقاطع حادثتين

K

Kurtosis	التفرطح

L

Law of total probability	قانون الاحتمال الكلي
Least square method	طريقة المربعات الصغرى
Level of significance	مستوي المعنوية
Linear correlation coefficient of Pearson	معامل الانحدار الخطي لبيرسون

M

Mathematical definition of probability	التعريف الرياضي للاحتمال
Mathematical expectation	التوقع الرياضي
Measure of dispersion	مقاييس التشتت
Measures of central f tendency	مقاييس النزعة المركزية
Median	وسيط
Mode	منوال
Mutually exclusive events	الحوادث المتنافية

N

Negative binomial distribution	توزيع ذي الحدين السالب
Negative skewed	ملتوياً ناحية اليسار

Non – ordered samples	العينات غير المرتبة
Normal distribution	التوزيع الطبيعي
Null hypothesis	فرض العدم

O

Observations	مشاهدات
One tailed	ذو طرف واحد
Ordered sample	العينات المرتبة

P

Packages	حزم
Parameters	معالم
Percentages	مئينات
Permutation	التباديل
Pie Chart	الدائرة البيانية
Poisson distribution	توزيع بواسون
Population	مجتمع
Population parameters	معالم المجتمع
Positively skewed	ملتوياً ناحية اليمين
Probability	الاحتمالات
Probability distribution	التوزيع الاحتمالي

Q

Qualitative data	بيانات وصفية
Qualitative variable	متغير وصفي
Quantitative data	بيانات كمية
Quantitative variable	متغير كمي

Quartiles	ربيعات

R

Random event	حادثة عشوائية
Random experiment	تجربة عشوائية
Random variable	متغير عشوائي
Random variable description	وصف المتغير العشوائي
Range	مدي
Rank coefficient of correlation for Spearman	معامل ارتباط الرتب لسبيرمان
Real	حقيقي
Rejection region	منطقة الرفض
Relation between events	العلاقة بين الحوادث
Relative	نسبي
Relative dispersion	التشتت النسبي
Relative frequency histogram	مدرج تكراري نسبي
Row data	البيانات المفردة

S

Sample	عينة
Sample space	فراغ العينة
Sampling distribution	توزيعات المعاينة
Sampling distribution of the mean	توزيع المعاينة للوسط
Sampling distribution of the sample proportion	التوزيع الاحتمالي للنسبة في العينة
Sampling distribution of the sample variance	التوزيع الاحتمالي لتباين العينة
Sampling with replacement	السحب بإرجاع
Sampling without replacement	السحب بدون إرجاع

Scatter diagram	شكل الانتشار
Semi interquartile range	نصف المدى الربيعي
Simple event	حادثة بسيطة
Simple linear regression	الانحدار الخطي البسيط
Simple random sample	العينة العشوائية البسيطة
Size of rejection region	حجم منطقة الرفض
Skewness	الالتواء
Some probability theorems	بعض نظريات الاحتمالات
Standard deviation	الانحراف المعياري
Standard Normal distribution	التوزيع الطبيعي القياسي
Standard sample	العينة المعيارية
Standard unit	الوحدات القياسية
Statistical inference	الاستدلال الإحصائي
Statistical packages	حزم إحصائية
Statistical testing of Hypotheses	اختبارات الفروض الإحصائية
Statistics	إحصاء
Stratified random sample	العينة الطبقية
Student`s t distribution	توزيع ت
Sub event	حادثة جزئية
Sure event	حادثة مؤكده
Symmetric	متماثل
Systematic sample	العينة المنتظمة

T

The multiplication rule	قاعدة ضرب الاحتمالات

Theorem	نظرية
Two tailed	ذو طرفين
Type I error	خطأ من النوع الأول
Type II error	خطأ من النوع الثاني

U

Uniform distribution	التوزيع المنتظم
Union of two events	اتحاد حادثتين

V

Variability	تغير
Variable	متغير
Variance	التباين

W

Weighted mean	متوسط موزون

كشاف المصطلحات

INDEX

ج

ح

نبذة عن المؤلف

عبدالرحمن بن إبراهيم محمد الخضيري

- ولد عام ١٣٨٨ هـ.
- حصل على بكالوريوس العلوم في بحوث العمليات عام ١٤١٥هـ من جامعة الملك سعود.
- حصل على ماجستير العلوم في بحوث العمليات عام ١٤٢١هـ من جامعة لانكاستر — المملكة المتحدة(بريطانيا).
- حصل على دكتوراه فلسفة العلوم في الرياضيات عام ١٤٢٥هـ من جامعة بيرمنجهام — المملكة المتحدة(بريطانيا).
- يعمل حالياً في قسم الإحصاء وبحوث العمليات بكلية العلوم - جامعة الملك سعود.
- يعمل حاليا وكيلاً لجامعة سلمان بن عبدالعزيز.
- عمل عميداً لكلية العلوم والدراسات الإنسانية — جامعة الخرج (جامعة سلمان بن عبدالعزيز).
- عمل عميداً مكلفاً لعمادة القبول والتسجيل بجامعة الخرج(جامعة سلمان بن عبدالعزيز).
- عمل عميداً لكلية العلوم بالخرج.
- عمل وكيلاً لكلية العلوم بالخرج.
- عمل رئيساً لقسم الرياضيات بكلية العلوم بالخرج.
- شارك في تصميم العديد من البرامج التدريبية الخاصة بالإحصاء وبحوث العمليات وتطبيقاتها العملية.
- نشر العديد من الأبحاث العلمية في مجلات علمية محكمة وفى مؤتمرات علمية دولية ومحلية.
- قام بتحكيم عدد من الأبحاث والمشاريع البحثية.
- عضو الجمعية السعودية للعلوم الإحصائية.
- عضو جمعية بحوث العمليات البريطانية.
- عضو جمعية بحوث العمليات الأمريكية.

نبذة عن المؤلف

عبدالفتاح مصطفي محمد السيد

- ولد عام ١٩٧٥م.
- حصل على بكالوريوس العلوم في الرياضيات تخصص الإحصاء وعلوم الحاسب عام ١٩٩٧م من جامعة المنصورة - مصر.
- حصل على ماجستير العلوم في الرياضيات تخصص الإحصاء وعلوم الحاسب عام ٢٠٠٢م من جامعة المنصورة – مصر.
- حصل على دكتوراه فلسفة العلوم في الرياضيات تخصص الإحصاء وعلوم الحاسب عام ٢٠٠٥م من جامعة المنصورة - مصر.
- يعمل مدرساً (أستاذ مساعد) في قسم الرياضيات – كلية العلوم – جامعة المنصورة – مصر.
- يعمل حالياً أستاذاً مشاركاً بكلية العلوم والدراسات الإنسانية – جامعة سلمان بن عبدالعزيز.
- رئيس وحدة التقنية بوحدة التطوير والجودة بكلية العلوم والدراسات الإنسانية – جامعة سلمان بن عبدالعزيز.
- رئيس وحدة الاحصاء بوحدة التطوير والجودة بكلية العلوم والدراسات الإنسانية – جامعة سلمان بن عبدالعزيز.
- عمل أستاذاً مساعداً بكلية العلوم والدراسات الإنسانية بالخرج – جامعة الملك سعود.
- شارك في تدريب طلاب الإحصاء وعلوم الحاسب بالمنصورة – مصر، في العديد من البرامج التدريبية الخاصة بالإحصاء وتطبيقاتها العملية.
- نشر العديد من الأبحاث العلمية في مجلات علمية محكمة وفى مؤتمرات علمية دولية ومحلية.
- قام بتحكيم عدد من الأبحاث والمشاريع العلمية والكتب.
- عضو جمعية الرياضيات المصرية.
- عضو في العديد من اللجان الداخلية بقسم الرياضيات –كلية العلوم والدراسات الإنسانية – جامعة سلمان بن عبدالعزيز.
- عضو لجنة تطوير الخطط الدراسية ومقررات الإحصاء بقسم الرياضيات – كلية العلوم – جامعة المنصورة – مصر.
- عضو لجنة الجودة والاعتماد الأكاديمي في قسم الرياضيات –كلية العلوم – جامعة المنصورة – مصر.

ملاحظات

ملاحظات

ملاحظات...

ملاحظات

مطبعة الحميضي ت: 2130130 الرياض

د. عبد الرحمن بن إبراهيم محمد الخضيري

- ولد عام ١٣٨٨هـ.
- حصل على بكالوريوس العلوم في بحوث العمليات عام ١٤١٥هـ من جامعة الملك سعود.
- حصل على ماجستير العلوم في بحوث العمليات عام ١٤٢١ هـ من جامعة لانكاستر - في لانكاستر - المملكة المتحدة (بريطانيا).
- حصل على دكتوراه الفلسفة في علوم الإحصاء والرياضيات عام ١٤٢٥هـ من جامعة بيرمنجهام - المملكة المتحدة (بريطانيا).
- يعمل حالياً في قسم الإحصاء وبحوث العمليات بكلية العلوم - جامعة الملك سعود.
- يعمل حالياً وكيلاً لجامعة سلمان بن عبد العزيز.
- عمل عميداً لكلية العلوم والدراسات الإنسانية بجامعة الخرج (جامعة سلمان بن عبد العزيز).
- عمل عميداً مكلفاً لعمادة القبول والتسجيل بجامعة الخرج (جامعة سلمان بن عبد العزيز).
- عمل عميداً لكلية العلوم بالخرج.
- عمل وكيلاً لكلية العلوم بالخرج.
- عمل رئيساً لقسم الرياضيات بكلية العلوم بالخرج.
- شارك في تصميم العديد من البرامج التدريبية الخاصة بالإحصاء وبحوث العمليات وتطبيقاتها العملية.
- نشر العديد من الأبحاث العلمية في مجلات علمية محكمة وفي مؤتمرات علمية دولية ومحلية.
- قام بتحكيم عدد من الكتب والأبحاث والمشاريع البحثية.
- عضو الجمعية السعودية للعلوم الإحصائية.
- عضو جمعية بحوث العمليات البريطانية.
- عضو جمعية بحوث العمليات الأمريكية.

د. عبد الفتاح مصطفى محمد السيد

- ولد عام ١٩٧٥م.
- حصل على بكالوريوس العلوم في الإحصاء وعلوم الحاسب عام ١٩٩٧م من جامعة المنصورة - مصر.
- حصل على ماجستير العلوم في الإحصاء وعلوم الحاسب عام ٢٠٠٢م من جامعة المنصورة - مصر.
- حصل على دكتوراه فلسفة العلوم في الإحصاء وعلوم الحاسب عام ٢٠٠٥م من جامعة المنصورة - مصر.
- يعمل مدرساً (استاذ مساعد) في قسم الرياضيات - كلية العلوم - جامعة المنصورة - مصر.
- يعمل حالياً أستاذاً مشاركاً بكلية العلوم والدراسات الإنسانية - جامعة سلمان بن عبد العزيز.
- رئيس وحدة التقنية بوحدة التطوير والجودة بكلية العلوم والدراسات الإنسانية - جامعة سلمان بن عبد العزيز.
- رئيس وحدة الإحصاء بوحدة التطوير والجودة بكلية العلوم والدراسات الإنسانية - جامعة سلمان بن عبد العزيز.
- عمل أستاذاً مساعداً بكلية العلوم والدراسات الإنسانية بالخرج - جامعة الملك سعود.

قام بالمشاركة في تدريب طلاب الإحصاء وعلوم الحاسب بالمنصورة - مصر، في العديد من البرامج التدريبية الخاصة بالإحصاء وتطبيقاتها العملية.

- قام بنشر العديد من الأبحاث العلمية في مجلات علمية محكمة وفي مؤتمرات علمية دولية ومحلية.
- قام بتحكيم عدد من الأبحاث والمشاريع العلمية والكتب.
- عضو جمعية الرياضيات المصرية.
- عضو في العديد من اللجان الداخلية بقسم الرياضيات - كلية العلوم والدراسات الإنسانية - جامعة سلمان بن عبد العزيز.
- عضو لجنة تطوير الخطط الدراسية ومقررات الإحصاء بقسم الرياضيات - كلية العلوم - جامعة المنصورة - مصر.
- عضو لجنة الجودة والاعتماد الأكاديمي في قسم الرياضيات - كلية العلوم - جامعة المنصورة - مصر.

ردمك: ٦-٠-٩٠٤٤٤-٦٠٣-٩٧٨
ISBN: 978 - 603 - 90444 - 0 - 6

مطبعة الحميضي ت: 2130130 الرياض